Membrane-Active Protein Interactions with Phospholipid Bilayers

by

© *Mohammad Hassan Khatami*

A thesis submitted to the

School of Graduate Studies

in partial fulfilment of the

requirements for the degree of

Doctor of Philosophy

Department of *Physics and Physical Oceanography*

Memorial University of Newfoundland

September 2016

St. John's Newfoundland

Abstract

Membrane-active proteins are a class of proteins that interact with lipid membranes in the body. I study two kinds of membrane-active proteins, antimicrobial peptides (AMPs) and lung surfactant (LS) proteins. In the first part of my PhD project I did computer simulation studies with two AMPs, Gaduscidin-1 and -2 (GAD-1 and GAD-2). These peptides are histidine rich and thus expected to exhibit pH-dependent activity. In this work I have performed molecular dynamics (MD) simulations with the peptides in both histidine-charged and histidine-neutral forms, along with 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid molecules, employing GROMACS software and an OPLS-AA force field. My results show a high tendency for pairs of histidines to interact with pore regions in both histidine-charged and histidine-neutral simulations. This work is published in Biophysica et Biochimica Acta (BBA)-Biomembranes (2014).

In the second part of my PhD research I performed computational simulations on lung surfactant protein B (SP-B) interacting with lipid bilayer. SP-B is a hydrophobic protein with 79 residues, from the saposin superfamily. Because of the extreme hydrophobicity of SP-B, the experimental structure of the protein is unknown. Thus, I combined the Mini-B (a fragment of SP-B) experimental structure and homology modelling based on proteins in saposin family to construct my initial model of SP-B. I run MD (using OPLS-AA and PACE force fields) and replica-exchange MD (using PACE force field) simulations with GROMACS software. I modelled SP-B in open and bent (V-shaped) structures, placed within or near a POPC lipid bilayer. My results demonstrate energetically feasible structures for SP-B, in which salt bridges

play a significant role. My simulations provide hypotheses for how SP-B promotes the rearrangement of planar lipid bilayers. Part of this work has been accepted for publication in Biophysica et Biochimica Acta (BBA)-Biomembranes (2016).

In the third part of my project I employed solid state nuclear magnetic resonance (NMR) using ^{2}H, ^{31}P and ^{15}N experiments, to study SP-B interacting with mechanically oriented lipid bilayer. Here, I used full-length ^{15}N-labelled SP-B, which was recombinantly expressed in our lab, to find the orientation of protein with respect to the bilayer. In this part of my thesis, the final goal was to compare the experimental ^{15}N spectra with the spectra, predicted from the structures we got from computational simulations to help define the protein's structure. Since, I was not able to gain ^{15}N NMR signals in my SP-B in lipid bilayer experiments, I did not fulfill the final goal of this part of my project. However, I was able to predict ^{15}N NMR spectra of my computational SP-B structures. My NMR results indicate that more optimization needs to be done to modify our SP-B preparation protocol to 1) increase the yields of isotope-labelled protein and 2) increase the protein:lipid ratio when refolding into lipids. My simulated ^{15}N spectra indicate that uniform ^{15}N-labelling is unlikely to constrain SP-B's structure and topology very much and it will likely be necessary to use a more specifically labelled sample for these experiments.

Acknowledgements

I would like to express my sincere appreciation to my supervisors, Dr. Valerie Booth and Dr. Ivan Saika-Voivod, for their guidance, encouragement, belief, patience and supervision, without which my PhD program would not be possible. I am grateful for the chance they gave me by accepting me as a PhD student in their group, and their great financial support. I would specially like to sincerely appreciate Professor Dr. Michael Morrow, who generously taught me theories of solid-state NMR and how to work with his NMR facilities. I would like to appreciate my advisory committee members, Dr. Anand Yethiraj and Dr. Christopher Rowley for their insightful discussions during my PhD program.

I specially thank Donna Jackman for her valuable discussions and her help when working in the lab. I appreciate her for preparing (and teaching how to recombinantly express) SP-B protein samples. I would also like to thank Dr. Celine Schneider for teaching me how to work with 600 BRUKER NMR instrument.

I would like to express my appreciation to Compute Canada, ACEnet and the SciNet HPC Consortium, for providing computational facilities. A part of my computations were performed on the ACEnet, which is the regional high performance computing consortium for universities in Atlantic Canada, which is funded by the Canada Foundation for Innovation (CFI), the Atlantic Canada Opportunities Agency (ACOA), and the provinces of Newfoundland and Labrador, Nova Scotia, and New Brunswick. Another part of my computations were performed on the gpc supercomputer at the SciNet HPC Consortium. SciNet is funded by: the Canada Foundation for Innovation under the auspices of Compute Canada; the Government of Ontario;

Ontario Research Fund - Research Excellence; and the University of Toronto.

I would like to also express my appreciation to all my friends at Memorial University, especially, Ahmad Al-Modallal, Taimaz Bahadory, Mohammad Azargon, Payam Bagheri, Saeid Eskandari, Martin LeBlanc, Gagandeep Sandhu, Mark McDonald, Tadiwos Asrat, Ashkan Rahmani, Nathan Agnew and Marek Bromberek, for their support and discussions. I would like to thank Fred Perry, Donna Coombs, Maureen Wade, Daphne Corbett and Joy Simons in Physics department for their help during my studies.

Without financial support I would not have been able to fulfill my studies. Here, I would like acknowledge the financial support through the ACEnet Graduate Fellowship, my supervisors and Memorial University. I would also like to acknowledge the financial support from school of graduate studies for the Dean's Doctoral Award.

I would like to thank my lovely parents and my great brothers for supporting me. Finally, my special thanks to Parisa, my wife and love of my life, without whom I would not be able to finish my studies.

— Mohammad Hassan Khatami

This work is dedicated to the love of my life, Parisa.

Contents

[1]This chapter is published as an article in Biophysica et Biochimica Acta (BBA)-Biomembranes (2014).

[2]A manuscript based on the all-atom simulations in this chapter has been accepted for publication in Biophysica et Biochimica Acta (BBA)-Biomembranes (2016).

[3]SP-B preparation and purification were carried out primarily by Ms. Donna Jackman, Booth lab.

List of Abbreviations and Symbols

^{1}H ..... Hydrogen

^{2}H ..... Deuterium

^{15}N ..... Nitrogen

^{31}P ..... Phosphorus

AA ..... Amino Acid

AMP ..... Antimicrobial Peptide

BLAST ..... Basic Local Alignment Search Tool

CG ..... Coarse-Grained

CH_2 ..... Methylene

CH_3 ..... Methyl

CP ..... Cross-Polarization

C-CART ..... Centre for Chemical Analysis, Research and Training

C:M ..... Chloroform:Methanol

DPPC 1,2-dipalmitoyl-sn-glycero-3-phosphocholine

DSSP ..... Define Secondary Structure of Proteins

GAD ..... Gaduscidin

GROMACS ..... GROningen MAchine for Chemical Simulations

H-bond ..... Hydrogen-bond

LJ ..... Lennard-Jones

LS ..... Lung Surfactant

MD ..... Molecular Dynamics

NH_2– ..... Amino group

NMR ..... Nuclear Magnetic Resonance

OPLS ..... Optimized Potentials for Liquid Simulations

OPLS-AA ..... Optimized Potentials for Liquid Simulations-All Atom

PC ..... Phosphatidylcholines

PDB ..... Protein Data Bank

PE ..... Phosphatidylethanolamine

PG ..... Phosphatidylglycerol

POPC 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine

POPG 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol

ppm ..... Part Per Million

REMD ..... Replica Exchange Molecular Dynamics

Rg ..... Radius of gyration

RMSD ..... Root mean square deviation

SDS-PAGE ..... sodium dodecyl sulfate polyacrylamide gel electrophoresis

SP-A ..... Surfactant Protein A

SP-B ..... Surfactant Protein B

SP-C ..... Surfactant Protein C

SP-D ..... Surfactant Protein D

SPV ..... Swiss PDB Viewer

STRIDE ..... STRuctural IDEntification

VMD ..... Visual Molecular Dynamics

–COOH ..... Carboxyl group

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

In this thesis, I present computational and experimental studies on two types of membrane active proteins. In this present chapter, I give an overall introduction to the concepts and methods I use in my thesis. In Chapter 2, I present the computational study I carried out on two antimicrobial peptide paralogues, GAD-1 and GAD-2, which are derived from cod fish. The work presented in Chapter 2 has been published as an article in BBA-Biomembranes [1]. In Chapter 3, I describe my computational study of the lung surfactant protein, SP-B, interacting with a lipid bilayer. Part of this chapter has been accepted for publication in BBA-Biomembranes [2]. The goal of Chapter 4 is to study experimentally with solid state NMR the same SP-B and bilayer system as in Chapter 3. We present predictions of solid state NMR spectra, based on structures obtained in Chapter 3, and the experimental efforts to obtain the NMR spectra.

1.2 Proteins and Lipids

Proteins are biological macromolecular polymers composed of monomeric units called amino acids **(Figure 1.1)**. There are 20 different amino acids. All of the proteins in living cells are made of linear sequences of these amino acids. Each amino acid is made up of a C_α carbon, which acts as a connector between an amino group (NH_2-), a carboxyl group ($-COOH$), a hydrogen atom and an R group, or side chain, which distinguishes the different amino acids. These amino acids polymerize through the formation of a peptide bond that links the N of the amino group to the C of the carboxyl group while liberating H_2O. As a result, one end of the protein (the N-terminus) maintains its amino group, while the other end (the C-terminus) maintains its carboxyl group, unless of course further chemical alteration of the ends takes place. Traditionally, the amino acid sequence of a protein is written out from left to right, starting with the N-terminus and ending with the C-termus [3]. In general, if the number of amino acid residues in the sequence is greater than 2 but less than 50, it is called a peptide, and if the number of residues exceeds 50, it is called a protein.

Along with proteins, lipids are an important class of molecule in cells, particularly in membranes. Lipids tend to be soluble in non-polar solvents and are water insoluble [3, 4]. They contain two parts, a hydrophilic head group and hydrocarbon tail, which is hydrophobic. Depending on the particular lipid, the head group can be neutral (containing no charge), positively charged, negatively charged or even zwitterionic (containing both positively and negatively charged parts). In most lipids, the tail contains one or more hydrocarbon chains of potentially different length, but some lipids, such as cholesterol, can have more complex hydrocarbon structures. The am-

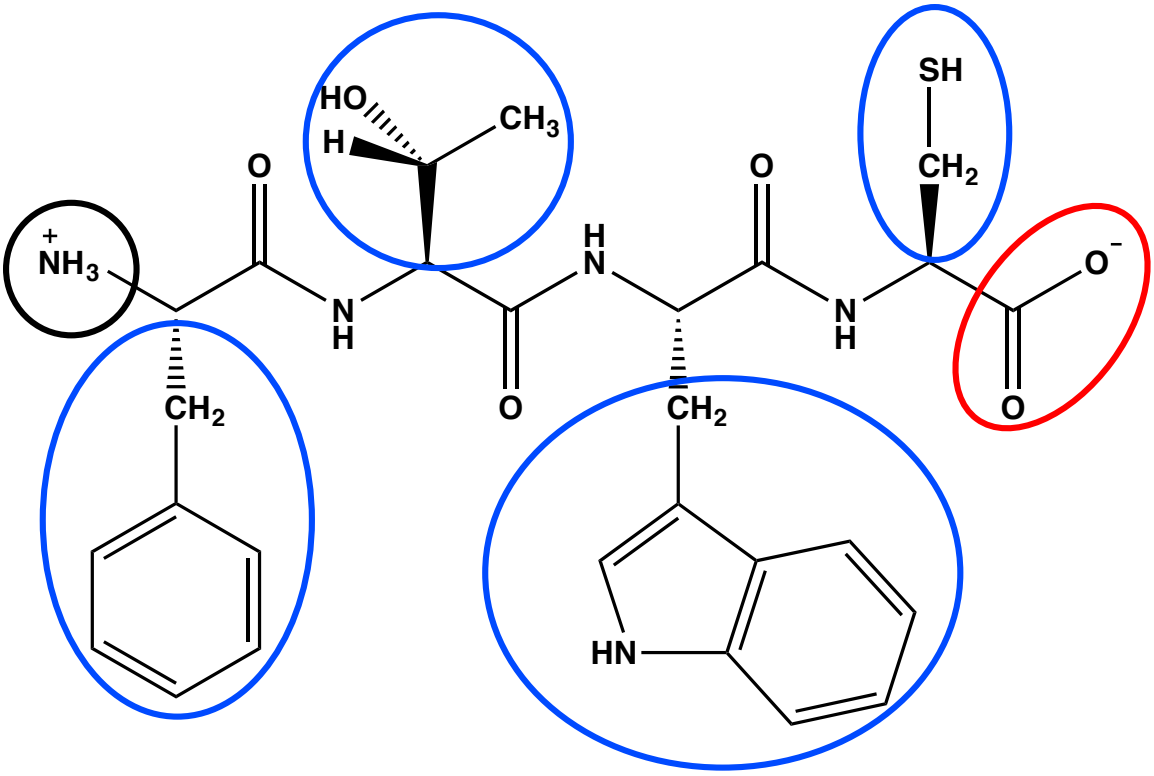

Figure 1.1: The schematic structure of a peptide containing four different amino acids. Here we can see the N- and C-terminals of the peptide, the backbone and residues. The N-terminus is circled in black, the C-terminus is circled in red, the side chains (residues) are circled in blue and the rest of the structure is the backbone.

phipathic properties of lipids help them to aggregate and form biological membranes.

1.3 Membrane-active Proteins

The outer layer of the cell is covered by a membrane which is composed of lipid molecules and proteins. The lipid part of the membrane is responsible for the shape and the structure of the membrane, while the proteins control the communications between the inner part of the cell and the regions outside the cell, i.e. signalling and

channeling across the membrane [5]. While membrane proteins enhance the functionality of the membrane, there are other types of proteins, like antimicrobial peptides (AMPs), that are termed "membane-active" and can interact with the membrane of target cells to kill them. In addition to cell membranes, there are also other lipid membranes of biophysical interest. Lung surfactant (LS), a mixture of lipids and proteins, that allows us to breathe by lowering the surface tension at the air-water interface within the lung.

As of 10th of April 2016, there are approximately 105 000 protein structures entered in the Protein Data Bank (PDB) [14], including those for approximately 1901 membrane proteins. Among these 1901, only 607 of them are unique proteins [15]. The main problem with membrane proteins is that they are intimately associated with membranes, and hence technically difficult to determine atomic resolution structures for them [5].

1.4 Lipid Bilayers

Lipid molecules are composed of two parts, a hydrophilic head group and a hydrophobic tail. When placed in water, these molecules, depending on their properties, generate specific structures to satisfy their hydrophobicity (**Figure 1.2**). Depending on the concentration and shape of the lipid molecules (the size of lipid headgroup, the length of lipid tail and their number), they can self-assemble into small structures, like micelles, or larger structures, like vesicles and lipid bilayers. Depending on the experimental methods employed, researchers study membrane-active proteins interacting with different lipid structures [7, 27, 29, 30]. For example, so-called solution

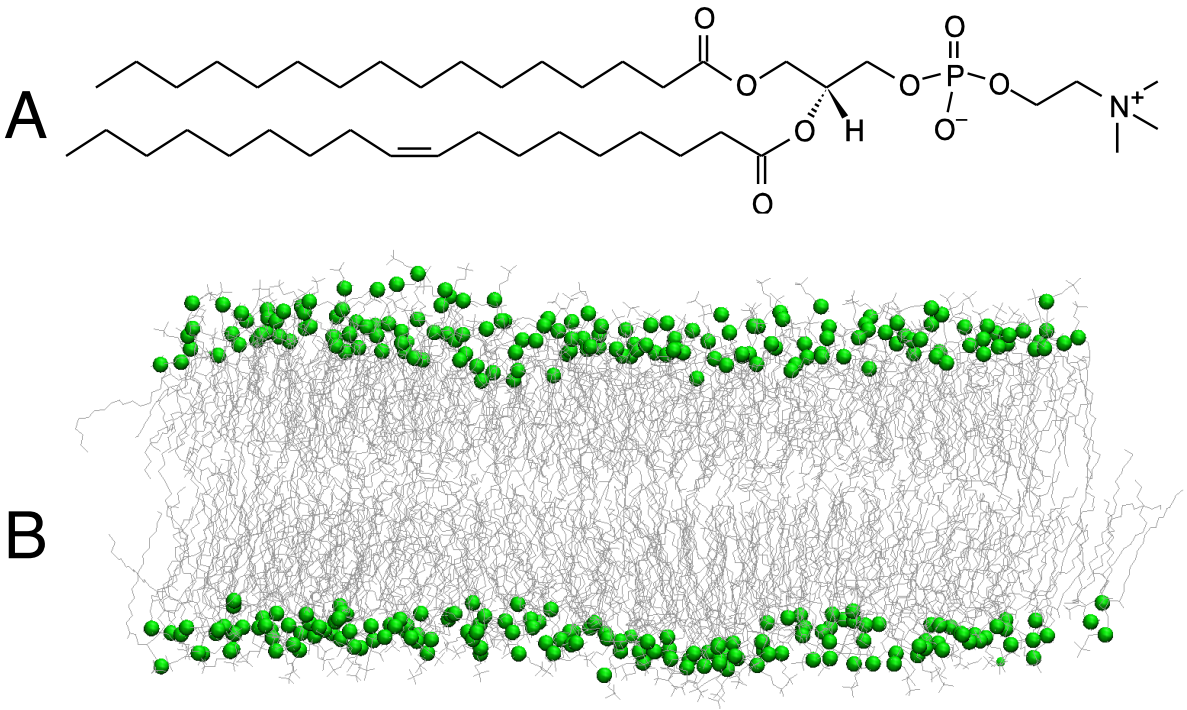

Figure 1.2: The schematic structure of POPC lipid molecule (top) and assembled lipid bilayer of 128 POPC lipid molecules (bottom). The green spheres represents the headgroup part of the POPC and the gray lines represents the acyl-chains of the lipids.

NMR is often done with proteins interacting with micelles. However, the chief goal is to determine structure and function of the protein while it is in its physiologically relevant environment.

In biological settings, membranes consist of mixtures of different kinds of lipids. A model lipid membrane chosen for a protein study, whether in simulation or experiment, should ideally represent the essential features of the biological system, be technically expedient for carrying out the study, and be as simple as possible. Here I provide some background motivating my choice of lipid.

Bacterial membranes mainly consist of lipids with phosphatidylglycerol (PG) and phosphatidylethanolamine (PE) headgroups, which are negatively charged and zwitterionic, respectively [10]. There is a variety of acyl chains encountered in bacterial membranes: saturated, unsaturated, single-chained, double-chained, short, long etc. POPC's zwitterionic headgroup is representative of headgroups in bacterial membranes and its acyl chains are not uncharacteristic of the same.

In lung surfactant, 80% of lipids are zwitterionic phosphocholine (PC) lipids, and half of these are 1,2-dihexadecanoyl-sn-glycero-3-phosphocholine (DPPC) [11] (40% of lipid content). Another major component of lung surfactant lipid is 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) [13]. In addition to PC lipids, there are 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol (POPG) lipids in lung surfactant [13]. POPC and POPG have the same two acyl-chains, a saturated chain with 16 carbons and an unsaturated chain with 18 carbons (**Figure 1.2-A**), but in DPPC both of the acyl-chains are saturated and have 16 carbon atoms. The saturated acyl-chains makes the DPPC lipid highly ordered and tightly packed as a bilayer, which causes its gel-to-liquid transition temperature to be rather high at 42°C, rendering it immobile (i.e. lipids do not diffuse) at body temperatures. Thus, POPC is a good model lipid in the context of lung surfactant studies, since it combines a common headgroup (PC) and common acyl chain (PO), and is mobile.

I use POPC for all the studies in this thesis: simulation of GAD-1 and GAD-2 AMPs, simulations of lung surfactant protein SP-B and solid state NMR experiments on SP-B. POPC is a commonly used model membrane in the context of simulations of protein-lipid interactions, since the force fields used to describe it reproduce experimental results with reasonable accuracy. In the solid state NMR experiments,

the fact that POPC is zwitterionic makes it significantly easier to prepare oriented samples using mica plates.

1.5 AMPs

Antimicrobial peptides have a key role in the immune systems of many different organisms, from humans to fish to insects [16–18]. In general, these peptides are amphipathic (having hydrophilic and hydrophobic faces) and cationic (positively charged) in nature [19], and thus, well suited to interactions with phospholipid bilayers. The cationic nature of AMPs helps them to interact with charged head groups of membranes while their amphipathicity helps them to interact with fatty acyl-chains. The positive charge of the AMPs facilitates their interaction with anionic lipids, such as those in bacterial membranes, but in high concentrations they are also toxic to mammalian cell membranes, even though these membranes are less anionic, in that they contain a smaller percentage of negatively charged residues. There are other studies, where in addition to the role of AMPs in the innate immune system, they are scrutinized for their potential to be used as a treatment for drug resistant infections [24–26] and cancer [27, 28, 42].

Nguyen et al. [20] recently reviewed 20 different ways of how AMPs interact with the membrane of their target bacteria **(Figure 1.3)**. In their work, they give account of the importance of secondary structure for how AMPs interact with their target, which can be either the membrane itself or elements in the interior of the cell, such as DNA. Extended AMPs, which do not have a regular secondary structure, usually interact with lipid intracellular targets and do not disturb the lipid membrane [20].

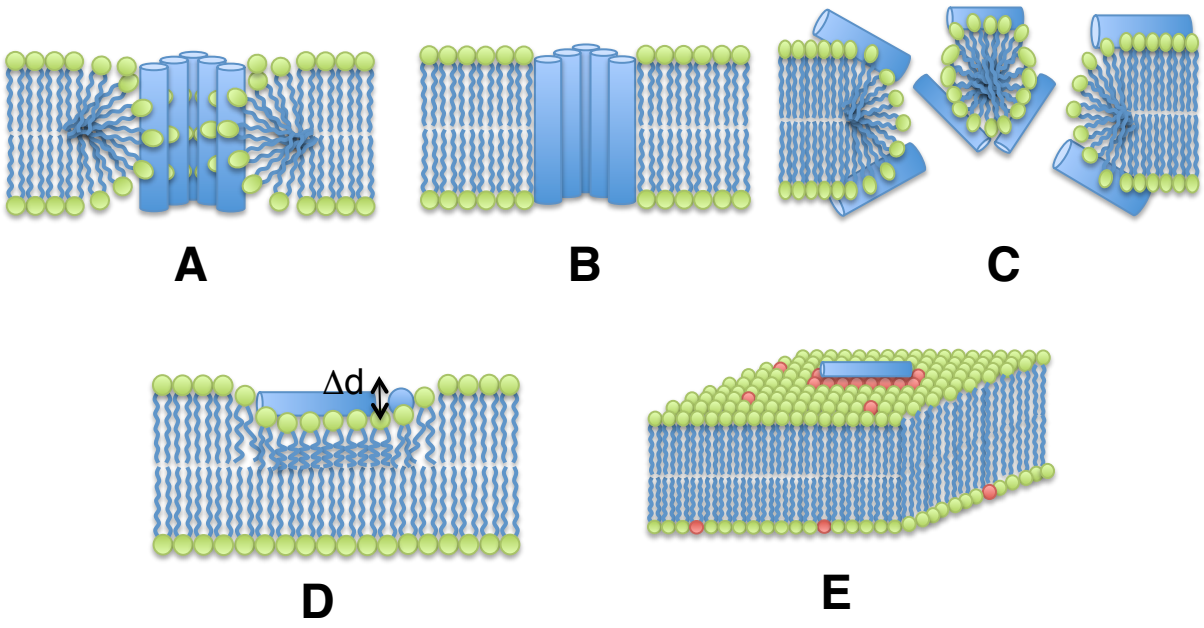

Figure 1.3: Example ways of α-helical AMPs interacting with lipid bilayers. A) Toroidal pore, B) Barrel stave, C) Carpet model, D) Membrane thinning and E) Charged lipids clustering. Figure is inspired from Ref. 20.

Figure 1.3 shows examples of how α-helical AMPs can interact with lipid bilayers, including forming pores. The length of α-helical AMPs is usually long enough to allow simultaneous interaction with both leaflets of a lipid membrane. **Figure 1.3-A** shows a toroidal pore, for which lipid headgroups are interspersed between the peptides as they form a curved pore lining. Depending on the amphipathicity of the α-helices, they can alternatively produce a so-called barrel stave, for which the hydrophilic parts of the AMPs face towards the water and the hydrophobic parts face towards the lipid acyl-chains **(Figure 1.3-B)**. In high concentrations, AMPs can completely disrupt the membrane, making micelle-like lipid structures and leaving big pores behind in the membrane **(Figure 1.3-C)**. α-helical AMPs can also induce membrane

thinning through localization on the bilayer surface **(Figure 1.3-D)**, which may help initiate pore formation. Membrane thickening is also possible. The AMPs can change the distribution of charged lipids by attracting anionic lipids **(Figure 1.3-E)**. The resulting clustering of anionic lipids may in turn affect the stability of the membrane.

As my simulations will show, GAD AMPs exhibit heterogenous structure with varying degrees of helicity, and thus do not necessarily follow the classical α-helical AMP picture. The different ways in which the polymorphic or disordered characteristics of the GAD AMPs allow them to interact with planar and pore-like portions of lipid bilayers is the subject of Chapter 2. The work described in Chapter 2 has been published in Biochimica et Biophysica Acta (BBA) - Biomembranes.

1.6 Lung Surfactant

Lung surfactant, a complex mixture of lipids and proteins, is found at the surface of the water film that lines the air sacs (alveoli) of the lungs. Lung surfactant prevents the lungs from collapsing by decreasing the surface tension at the air-water interface within alveoli, and thus allows us to breathe. Its action greatly decreases the effort required of breathing muscles, i.e. the diaphragm and intercostal muscles of the rib cage [12].

To help understand the physics behind breathing, we present a simple model of the lung in **Figure 1.4-A**, where we represent the lung as a single spherical alveolus and the thoracic cavity as a box around it. We need to introduce the Young-Laplace equation to help explain the role of LS. It relates pressures inside and outside a bubble-like object in mechanical equilibrium with the so-called Laplace pressure that

arises from surface tension,

$$\Delta P_{i,c} = P_i - P_c = \frac{2\gamma}{R}, \tag{1.1}$$

where P_i is the pressure inside the alveolus, P_c is the pressure outside the aveoulus (the so-called intrapleural pressure of the thoracic cavity), $\Delta P_{i,c}$ is the difference between the two, R is the radius of the air sac, which is approximately 100 μm in diameter [31], and γ is the surface tension. In equilibrium, γ is about 25 mN/m [11], which through **Equation 1.1** gives $2\gamma/R = 500$ N/m$^2 \simeq 4$ mmHg. Thus, P_i needs to be greater than P_c by this amount in order to balance the tendency of surface tension to shrink the alveolus. In actual fact, P_c is generally below atmospheric pressure (a gauge pressure of roughly -5 mmHg in equilibrium) [12], and this prevents the lung from collapsing. One should note that lung function requires the pressure inside the air sacs to be very near to atmospheric pressure [29].

In a typical breathing cycle, contraction of the breathing muscles during inhalation increases the volume of the chest cavity, which decreases P_c from -5 mmHg to -10 mmHg. In response, the lung expands and P_i becomes slightly lower than atmospheric pressure. This lower value of P_i allows air to flow into the lung. Here, we seem to have a contradiction: R has increased, implying that $2\gamma/R$ has decreased, and yet $\Delta P_{i,c}$ has increased. The answer to this puzzle is that γ, in fact, increases as R increases.

Of course this simple view neglects the complex dynamics of breathing. However, it does provide a rationale for why it is beneficial that γ should increase as R increases. Physically, γ increases because the surfactant area concentration decreases as an alveolus expands.

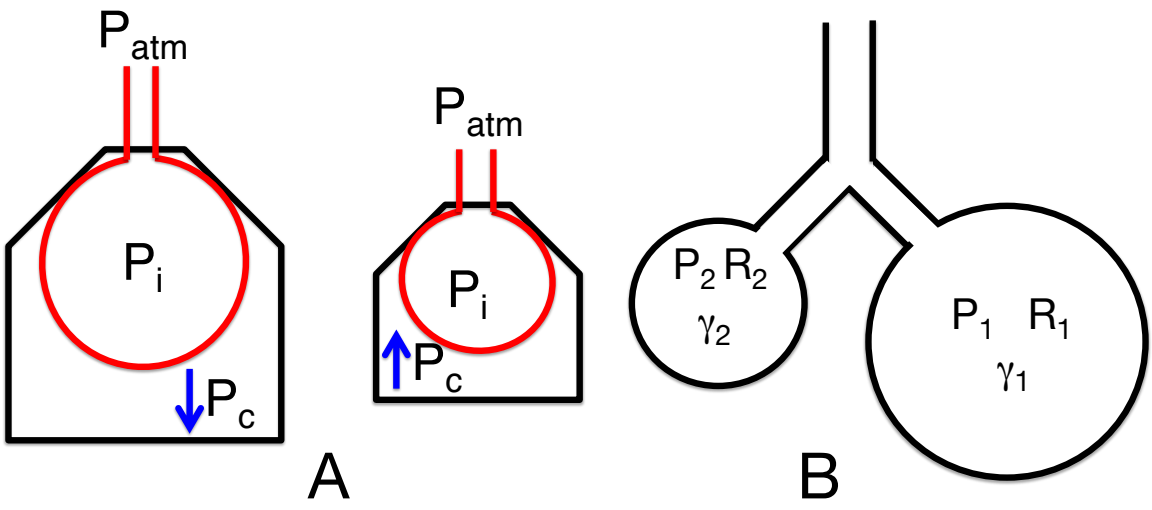

Figure 1.4: Schematics of lungs. A) Schematic steps in breathing, and B) Pair of different attached alveoli. The irregular hexagon represent the thoracic cavity and the circle represents an alveolus. P_{atm} represents atmospheric pressure, P_i represents the pressure inside the alveoli and P_c represents the intrapleural cavity pressure. During inhalation the volume of thoracic cavity increases, which decreases the intrapleural pressure (A-left), while during exhalation the volume of thoracic cavity decreases, which increases the intrapleural pressure (A-right).

During exhalation, the breathing muscles relax in order to decrease the volume of the chest cavity, which increases P_c up to a maximum of nearly atmospheric pressure. In response, the lung shrinks and the pressure inside the lung becomes slightly above atmospheric pressure (roughly +1 mmHg) [12]. This higher value of pressure inside the lung produces an air flow out of the lung. At the end of exhalation, $\Delta P_{i,c}$ has decreased through nearly 0. If the alveolus is to remain in mechanical equilibrium, and not collapse because of the Laplace pressure, γ must also approach zero. This

occurs in the shrinking alveolus as the surfactant area concentration increases.

To further appreciate the role of the surface tension's flexibility in alveoli, we should study the behaviour of two connected alveoli, schematically represented in **Figure 1.4-B**. Using **Equation 1.1**, we have:

$$\Delta P_{1,c} = P_1 - P_c = \frac{2\gamma_1}{R_1},$$

$$\Delta P_{2,c} = P_2 - P_c = \frac{2\gamma_2}{R_2}.$$

Assuming constant γ, $P_2 > P_1$ since R_2 is smaller than R_1, and hence the air inside the smaller alveolus should flow into the bigger one, causing the smaller one to collapse and the bigger one to possibly burst. However, if γ decreases with the decreasing R, then γ_2 could be nearly 0 while γ_1 has a non-zero, positive value. Then, based on **Equation 1.1**,

$$\Delta P_{2,c} = P_2 - P_c \simeq 0 \ \rightarrow \ P_c = P_2,$$

$$\Delta P_{1,c} = P_1 - P_c = P_1 - P_2 = \frac{2\gamma_1}{R_1} > 0.$$

Thus $P_1 > P_2$, and air flows from alveolus 1 into alveolus 2, which equilibrates the pressure inside the lungs.

There are four different types of surfactant proteins, SP-A, SP-B, SP-C and SP-D. SP-A and SP-D are hydrophilic proteins, while SP-B and SP-C are hydrophobic proteins. SP-B is an essential protein for life, without which we can not breathe [32]. Despite its importance, its 3D structure is unknown and therefore its mechanism of action is poorly understood.

The second part of my research is devoted to studying SP-B, both computationally and experimentally. Chapter 3 describes molecular dynamics and replica-exchange

molecular dynamics simulations of SP-B interacting with a POPC lipid bilayer. I use two force fields, an all-atom force field (OPLS-AA) and a potential that employs coarse-graining (PACE) in hopes of increasing the speed of simulations without losing essential physics. The parts related to the all-atom force field simulations of Chapter 3 are accepted to be published in Biochimica et Biophysica Acta (BBA) - Biomembranes. In Chapter 4, I present my solid state NMR experiments on SP-B in a mechanically oriented lipid bilayer, as well as predictions of NMR spectra from simulations.

1.7 Methods

1.7.1 MD Simulations

Molecular dynamics simulation is a computational technique to simulate N-body systems, e.g. biomolecular systems, and compute their equilibrium and transport properties [43]. In MD simulations we solve Newton's equations of motion for each particle in our system,

$$m_i \frac{\partial^2 r_i}{\partial t^2} = F_i, \quad i = 1, \cdots, N, \tag{1.2}$$

where r_i is the position of particle i, N is the number of particles in the system and the force and F_i is the net force on particle i.

Solving the coupled system of ordinary differential equations in **Equation 1.2** provides us with both the position and velocity of each particle at every time step, which generates the "trajectory" of system configurations that is useful for further structural analysis. To solve the equations of motion, we need to use an appropriate

numerical method, i.e. one that is reasonably accurate and computationally fast. The Verlet algorithm [44, 45] is often used as it produces reasonably accurate trajectories for short times and conserves energy at long times.

The potential energy governing the interactions between atoms in the system is specified by a so-called "force field". The differences between force fields emerge from different approaches used for approximating the essentially quantum mechanical interactions between atoms with classical pair potentials. A force field models both non-bonded interaction, i.e. Coulombic and van der Waals forces, and bonded interaction, i.e. angular, translational and torsional harmonic interactions. Each kind of interaction is described by a particular functional form and a set of parameters that characterize a particular atomic species. In our simulations we employ a version of the all-atom Optimized Potential for Liquid Simulations (OPLS-AA) force field [46, 47] adapted for POPC lipid molecules [48].

Additionally, depending on the stage of our simulation protocol, we run our systems under NVT (Canonical) or NPT (isothermal-isobaric) ensembles. For the NVT ensemble, the number of particles (N), volume of the system (V) and temperature of the system (T) are held constant. Similarly, under the NPT ensemble, the number of particles, pressure of the system (P) and temperature are held constant. To keep the pressure and the temperature constant, one uses pressure couplings (barostats) and temperature couplings (thermostats), respectively. The two main thermostats that are frequently being used in MD simulations are the Berendsen [49] and Nose-Hoover schemes [50,51]. The Berendsen thermostat is simpler, and works by rescaling velocities so that their average approaches what is expected from the equipartition theorem; it does not rigorously recover the canonical ensemble. In my simulations,

I use Nose-Hoover thermostat, in which the Hamiltonian of the system is extended by adding a coordinate that is coupled to the particle coordinates and the equation of motion of which depends on the temperature. This term works as an oscillatory damping term, which keeps the temperature of the system constant and ensures energy fluctuations follow those of the canonical ensemble [41]. The two main barostats being used in MD simulations are Berendsen algorithm [49] and Parrinello-Rahman approach [34, 35]. The Berendsen barostat rescales the simulation box (and particle coordinates commensurately) so that the desired pressure is approached. Here, I employ Parrinello-Rahman barostat, which theoretically provides true NPT ensemble, to run my simulation. This approach is similar to Nose-Hoover thermostat scheme, in which it uses oscillatory damping method to keep the pressure constant.

The advantage of using classical MD simulations is being able to determine the positions and velocities of all particles in the system for relatively long times. However, since we are using classical equations of motion, we can not simulate any strongly quantum mechanical behaviour, such as hydrogen atom motion or H-bond vibrations in our system. Thus, the modelling of certain intramolecular interactions is simplified by replacing vibrational degrees of freedom with constraints, e.g. water molecules are treated as rigid V-shaped objects decorated with interaction sites. Moreover, using classical force fields one obviously can not calculate electronic properties. There are other limitations in MD simulation arising from limitations in computing power. The simulation box size is usually limited to contain at most few million atoms because of the sheer number of calculations required for this number of atoms. The duration of simulations for large systems, i.e. more than hundreds of thousands particles, is usually limited to less than a few microseconds, which is short compared to the duration

of experiments.

One of the fastest packages for running MD simulations [55] is GROningen MAchine for Chemical Simulations (GROMACS), which is designed to work with different force fields. GROMACS is a free engine for classical MD simulations and energy minimizations of biological systems like lipids, proteins and nucleic acids in various environments, e.g. aqueous media. I use this package to run my simulations.

1.7.2 Force Fields

Each force field has two main parts: 1) equations of a particular mathematical form that describe potential energies needed to generate forces and equations of motion and 2) parameters within these equations that are tuned so that the physical properties of the simulated system are comparable with experimental results [41]. In GROMACS, the equations for potential energies are divided into those for non-bonded forces, and those for bonded forces.

Non-bonded forces, e.g. Lennard-Jones (LJ) and Coulomb forces, act between atoms that are either on different molecules or sufficiently far apart on the same molecule. For computational efficiency, interaction between non-bonded atoms are explicitly calculated only up to a finite cutoff distance. In GROMACS, these forces and potentials are pair-additive and centro-symmetric:

$$V(r_1, \cdots, r_N) = \sum_{i<j} V_{ij}(r_{ij}) \tag{1.3}$$

$$F_i = -\sum_j \frac{dV_{ij}(r_{ij})}{dr_{ij}} \frac{\mathbf{r_{ij}}}{r_{ij}} \tag{1.4}$$

V represents the potential energy of the system, V_{ij} represents the potential energy between two particles i and j, r_{ij} is the distance between particle i and particle j and $\mathbf{r_{ij}}$ represents the vector from particle j to particle i. Within OPLS-AA, the LJ potential is formulated as,

$$V_{ij}(r_{ij}) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], \tag{1.5}$$

where ϵ_{ij} and σ_{ij} are parameters. Each particle i has an associated σ_i and ϵ_i depending on its type. These parameter values are explicitly given in OPLS-AA. OPLS-AA then uses geometrical combination rules to find the values for LJ constants, $\sigma_{ij} = (\sigma_i\sigma_j)^{1/2}$ and $\epsilon_{ij} = (\epsilon_i\epsilon_j)^{1/2}$.

For charged particles in the system, the Coulomb potential describes their interaction,

$$V_c(r_{ij}) = \frac{f}{\varepsilon_r}\frac{q_i q_j e^2}{r_{ij}} = \frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{q_i q_j e^2}{r_{ij}}, \tag{1.6}$$

where e is electron charge, q_i is the elementary charge of atom i, ε_0 is the permitivity of free space and the relative dielectric constant ε_r is set to unity in our case [56].

Because of the long-range nature of Coulomb interaction, it is insufficient to simply employ a radial cutoff and therefore necessary to express the Coulomb energy of the system using Ewald summations [57] in order to calculate periodic electrostatic interactions. In the Ewald algorithm, the slow convergence of the energy with distance due to the $1/r$ form of the Coulomb interaction is dealt with by a clever mathematical reformulation of the system energy that involves reciprocal space sums. The Coulomb energy of a system of N particles embedded in an infinite array of periodic images is given by,

$$V = \frac{f}{2}\sum_{n_x}\sum_{n_y}\sum_{n_z}\sum_{i}^{N}\sum_{j}^{N}{}' \frac{q_i q_j}{r_{ij,n}} = V_{dir} + V_{rec} + V_0. \tag{1.7}$$

For a cubic simulation simulation box, $\vec{r}_{ij,n} = \vec{r}_{ij} + L(n_x, n_y, n_z)$ points from particle j to particle i in the periodic image of the simulation cell located at $L(n_x, n_y, n_z)$, and $r_{ij,n}$ is its magnitude. n_x, n_y and n_z are integers. The prime on the summation indicates that the terms with $i = j$ are omitted when $\mathbf{n} = (0, 0, 0)$. The direct space, reciprocal space and self-energy correction terms are given, respectively, by,

$$V_{dir} = \frac{f}{2} \sum_{ij}^{N} \sum_{n_x} \sum_{n_y} \sum_{n_z}{}' q_i q_j \frac{\mathrm{erfc}(\beta r_{ij,n})}{r_{ij,n}} \tag{1.8}$$

$$V_{rec} = \frac{f}{2\pi V_c} \sum_{ij}^{N} q_i q_j \sum_{m_x} \sum_{m_y} \sum_{m_z *} \frac{\exp\left(-(\pi \mathbf{m}/\beta)^2 + 2\pi i \mathbf{m}.(\mathbf{r_i} - \mathbf{r_j})\right)}{\mathbf{m}^2} \tag{1.9}$$

$$V_0 = -\frac{f\beta}{\sqrt{\pi}} \sum_{i}^{N} q_i^2, \tag{1.10}$$

where β is a parameter that controls how much weight is given to the direct potential and reciprocal potentials, $\mathrm{erfc}(x)$ is the complementary error function, V_c is the volume of the simulation cell, $\mathbf{m} = (m_x, m_y, m_z)$ is the reciprocal space index (integer) vector and the star indicates that the $\mathbf{m} = (0, 0, 0)$ term is omitted. While the energy is invariant with respect to β in the infinite limit, β is chosen so that a relatively short radial cutoff can be safely employed in the direct space sum while using as few terms as possible in the reciprocal space sum. In practice, especially for large systems, refined versions of this algorithm that make use of fast Fourier transforms, such as the particle mesh Ewald algorithm [58] are employed.

In the family of bonded interactions, there are the bond stretching (2-body), the bond angle (3-body) and the dihedral bond angle (4-body) interactions (**Figure 1.5**). A harmonic potential is introduced for the bond stretching interaction between two particles:

$$V_b(r_{ij}) = \frac{1}{2} k_{ij} (r_{ij} - b_{ij})^2, \tag{1.11}$$

where k is the spring constant and b is the bond length in equilibrium. In simulations, other terms are needed to calculate the energetics of the system more realistically and accurately. The bond angle interactions between triplets of particles can be modelled

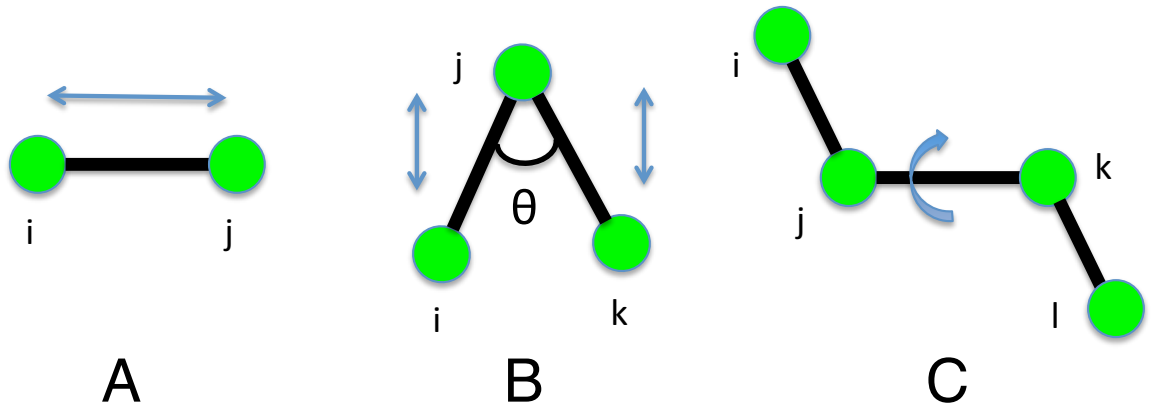

Figure 1.5: sketch of 3 types of bonded interactions. A) bond stretching, B) angle vibration and C) proper dihedral angle.

with simple harmonic potential,

$$V_a(\theta_{ijk}) = \frac{1}{2}k_{ijk}(\theta_{ijk} - \theta^0_{ijk})^2, \tag{1.12}$$

where i, j, k are the particle indices, k is the angle constant and θ^0 is the angle in equilibrium.

The proper dihedral angle torsional interaction is often described with a periodic function:

$$V_d(\phi_{ijkl}) = k(1 + \cos(n\phi - \phi_s)), \tag{1.13}$$

or Ryckaert-Bellemans function:

$$V_{rb}(\phi_{ijkl}) = \sum_{n=0}^{5} C_n(\cos(\psi))^n, \tag{1.14}$$

where ψ=ϕ–180°.

In OPLS-AA the torsional energy is expressed by a third-order Fourier expansion:

$$E_{torsion} = \sum_i \frac{V_{1,i}}{2}[1+\cos(\phi_i + f_{1,i})] + \frac{V_{2,i}}{2}[1-\cos(2\phi_i + f_{2,i})] + \frac{V_{3,i}}{2}[1+cos(3\phi_i + f_{3,i})],$$

where i represents each dihedral angle, ϕ_i is the dihedral angle, V_1, V_2, and V_3 are the coefficients in the Fourier series, and f_1, f_2, and f_3 are phase angles [56].

Depending on the simulation, one can use all-atom, united-atom or coarse grained force fields. An all-atom force field treats every atom explicitly and applies detailed equations of motions to all of the particles in the system. Simulations based on these detailed force fields are time consuming. OPLS-AA is the all-atom force field I use in my simulations. United-atom force fields treat all the particles in detail, except for the terminal methyl and methylene bridges, where C and H atoms are combined as a single bead. In coarse-grained force fields, the system is represented by beads, with each bead representing a group of particles, e.g. in the MARTINI [59] force field every four water molecules are represented with one bead, while for most other molecules constitute every four heavy atoms (e.g. C and/or N) is one bead, and H atoms are mainly ignored. Using coarse-grained force fields lowers the simulation time at the expense of losing details of the simulation.

1.7.3 REMD Simulations

Replica exchange molecular dynamics (REMD) is a method to help us sample the configurational space of the system more efficiently. In this method we simulate the same system at different temperatures (different replicas of our sample) at the same time. The idea behind this method is to randomly exchange the atomic configurations

between replicas. The probability of accepting an exchange of configurations between two replicas is given by [41],

$$P(i \leftrightarrow j) = \min\left(1, \exp\left[\left(\frac{1}{k_B T_i} - \frac{1}{k_B T_j}\right)(V_i - V_j)\right]\right), \tag{1.15}$$

where T_i and T_j are the temperatures and V_i and V_j are the potential energies of the two replicas i and j at the time of exchange. In this method, after exchanging the replicas, the velocities of particles are scaled by the factor of $(T_i/T_j)^{\pm 0.5}$. In each step, the REMD method employed by GROMACS performs exchanges only between one of the two neighbouring replicas. As an example if we have system 0, 1, 2 and 3 in the first step of exchange we have 0-1 and 2-3 exchange attempts, while in the second step we have 1-2 exchange attempt and this continues during the simulation. In between exchange attempts, the system evolves according to regular MD. The number of MD steps between exchange attempts is typically 100 steps. The number of exchange attempts should be calibrated in a way that the acceptance ratio to be roughly 20–30% to get higher efficiency [60].

To optimize the sampling of configurations between replicas over the range of temperature used, the replicas should be assigned temperatures according to an exponential distribution [61]:

$$T_i = T_0 \times \exp(k \times i), \tag{1.16}$$

where i is the number of replica, T_i is the temperature of each replica, T_0 is the temperature of the first replica and k is a constant value proportional to the total number of replicas, calculated using $k = \frac{1}{N_{rep}} \ln\left(\frac{T_f}{T_0}\right)$, where T_f is the temperature of the final replica and N_{rep} is the total number of replicas.

1.7.4 Solid State NMR

Solid-state nuclear magnetic resonance (SSNMR) is a spectroscopic technique for structural and dynamical studies of biological systems e.g. lipid bilayers and membrane proteins. Using different labelling isotopes like ^{15}N and ^{2}H on the target systems like proteins and lipid acyl chains and using the ^{31}P spectra [62] of phosphorus headgroups of lipid membranes can give us precious information about the structure and dynamics of the system. The Hamiltonian of the system can be written as [63]:

$$H = H_0 + H_{dip} + H_{quad} + H_{CS} + \cdots , \tag{1.17}$$

where H_0 is the Zeeman effect Hamiltonian, H_{dip} is the Hamiltonian of dipolar interactions, H_{quad} is the Hamiltonian of quadrupolar interactions and H_{CS} is the chemical shift due to local chemical shielding interactions in the system.

The ^{2}H-labelling of carbons in lipid acyl-chains gives us information about the order parameter of the bilayers based on quadrupolar interactions of deuterium (^{2}H) in a magnetic field. Here, the smaller the splitting, the lower the order parameter, which indicates greater fluctuation of the lipid chains. Usually, in the acyl-chain the methylene groups (CH_2) nearer to the lipid head group are more ordered and the order parameter decreases, when reaching to the methyl group (CH_3).

In addition to ^{2}H SSNMR experiments, the ^{31}P spectra of phosphorus atoms in lipid headgroups also provide us information on the orientation of lipid bilayers. This method uses the chemical shift anisotropy of ^{31}P.

Besides ^{2}H and ^{31}P SSNMR experiments, using ^{15}N chemical shifts can give us information about the structure and orientation of a protein in a lipid bilayer [30, 31].

Similar to ^{31}P experiments, here, we use the chemical shift anisotropy of ^{15}N. Since there are low amounts of the ^{15}N nuclei in experiments, using cross-polarization (CP) method we can enhance the signal to noise ratio in our experiments. In CP the polarization from frequent nuclei, e.g. ^{1}H, is transferred to dilute nuclei such as ^{15}N.

As a part of my research, I studied orientation and conformation of SP-B in an oriented bilayer (mechanically oriented bilayer supported on mica sheets) using solid state NMR. In chapter 4 I will give a general overview of how I employ NMR experiments in our project.

1.7.5 Homology Modelling

Homology modelling is a technique to predict the structure of a protein using homologous proteins, i.e., proteins with a common ancestor and hence similar amino acid sequences. Some early uses of this method can be found in papers by Perutz et al. [33, 34], where they discuss how to use the 3D structure of similar amino acid regions in horse oxyhaemoglobin and sperm whale myoglobin proteins to predict the yet unknown structure of haemoglobin. A more systematic study of the relationship between sequence identity and protein structure by Chothia et al. [35] indicated that to have a reasonably well predicted structure, the sequence identity of the protein and homologue should be $> 50\%$. Sequence identity refers to the fraction of positions in the amino acid sequences of two proteins that have the same amino acid residue. Through the years, homology modeling has improved. Recently, Sadowski et al. [36] introduced their method, in which the two protein could have only 35-40% sequence identity to produce a homology model. The SWISS-MODEL [37–40], which is a fre-

quent model to run homology modelling (http://swissmodel.expasy.org), employs the Sadowski et al. [36] method in their model.

1.8 Helicity Calculation

A hydrogen bond is an electrostatic interaction between a hydrogen attached to a highly electronegative atom, i.e. fluorine, oxygen or nitrogen, and a nearby highly electronegative atom. In the protein backbone, CO of amino acid i can make a hydrogen bond with NH of amino acid j. Depending on the patterns appearing in the repetition of hydrogen bonds (H-bonds) in the sequence, there are methods to identify the secondary structure of the protein based on the hydrogen bonding patterns.

Generally, proteins can take on α-helical, β-strands and sheets, 3_{10}-helix, π-helix, turn and random coil structures. Since our proteins mainly contain α-helical structures, we focus on this structure. For an α-helix (which is usually a right handed helix), the CO of amino acid i makes a hydrogen bond with the NH of amino acid i+4, and thus in each turn of the helix, there are 3.6 amino acids.

The main two methods of calculating the secondary structure of a protein are STRuctural IDEntification (STRIDE) [66] and the Definition of Secondary Structure of Protein (DSSP) [67, 68]. The DSSP method is the most common method in calculating secondary structure of a protein. This method uses the electrostatic hydrogen bond's energy between the CO and NH groups in the protein to define the structure. Here, the CO and NH groups have partial charges: + q_1 and – q_1 for CO and – q_2

and + q_2 for NH. The hydrogen bond energy in kcal/mol is represented as:

$$E = q_1 q_2 \left(\frac{1}{r(ON)} + \frac{1}{r(CH)} - \frac{1}{r(OH)} - \frac{1}{r(CN)} \right) \times f, \tag{1.18}$$

where q_1=0.42e and q_2=0.20e, with e being proton charge, r(ij) in Å is the distance of atom i from atom j and f=332 kcal Å/mol is a dimensional factor. The energy for a perfect H-bond is about -3 kcal/mol and a perfect N–O distance is d=2.9 Å. However, this method uses the cutoff energy of E $\leq$ -0.5 kcal/mol, which corresponds to d up to 5.2 Å when we have perfect alignment of CO and NH, i.e. θ (the angle between CO and NH)$\simeq$ 0, or d=2.9 Å when θ changes up to 63°.

For example, in this method, to have an α-helical structure in a fragment of protein, between i, i+3 AAs, we should have an H-bond between i-1, i+3 and i, i+4 (no additional information about i+1 and i+2 is needed). GROMACS uses this method to calculate the secondary structure of the protein.

The STRIDE model uses a more complex method to calculate the secondary structure of the protein. This method uses terms related to the properties of each specific amino acid. The H-bond energy is represented as:

$$E_{hb} = E_r \times E_t \times E_p, \tag{1.19}$$

where E_r is a function of distance, and E_t and E_p are functions of angles in the backbone (dihedral properties). In this method, if i, i+4 and i+1, i+5 satisfy the conditions for having hydrogen bonds, this method labels i+1, i+2, i+3 and i+4 as α-helical region (no additional information about i+2 and i+3 is needed). VMD program uses STRIDE as the default helicity calculator method.

Bibliography

[1] Khatami, M. H., Bromberek, M., Saika-Voivod, I., Booth, V. "Molecular dynamics simulations of histidine-containing cod antimicrobial peptide paralogs in self-assembled bilayers" (2014) BBA - Biomembranes 1838, 2778-2787.

[2] Khatami, M. H., Saika-Voivod, I., Booth, V. "All-atom Molecular Dynamics Simulations of Lung Surfactant Protein B: Structural Features of SP-B Promote Lipid Reorganization" (in press 2016) BBA - Biomembranes.

[3] Wallace, R. A., Sanders, G. P. and Ferl, R. J. "Biology: The science of life" (1990) (third edition), Harper Collins.

[4] Voet, D., Voet, J. G., and Pratt, C. W. "Fundamentals of Biochemistry " (2005) (second edition), John Wiley & Sons, Inc.

[5] Pebay-Peyroula, E. "Biophysical Analysis of Membrane Proteins: Investigating Structure and Function" (2008) WILEY-VCH.

[6] Bechinger, B., Resende, J. M., Aisenbrey, C. "The structural and topological analysis of membrane-associated polypeptides by oriented solid-state NMR spectroscopy: established concepts and novel developments" (2011), Biophys Chem, 153, 115-25.

[7] Sanders, C. R. and Sonnichsen, F. "Solution NMR of membrane proteins: practice and challenges" (2006) Magn Reson Chem, 44 Spec No, S24-40.

[8] Kandasamy, S. K., Lee, D. K., Nanga, R. P., Xu, J. Santos, J. S., Larson, R. G., Ramamoorthy, A. "Solid-state NMR and molecular dynamics simulations reveal the oligomeric ion-channels of TM2-GABA(A) stabilized by intermolecular hydrogen bonding" (2009), Biochim Biophys Acta, 1788, 686-95.

[9] Baoukina, S. and Tieleman, P. D. "Lung Surfactant Protein SP-B Promotes Formation of Bilayer Reservoirs from Monolayer and Lipid Transfer between the Interface and Subphase" (2011) Biophysical Journal, 100, 1678-1687.

[10] Epand, R. F., Savage, P. B., Epand, R. M. "Bacterial lipid composition and the antimicrobial efficacy of cationic steroid compounds (Ceragenins)" (2007) Biochimica et Biophysica Acta 1768 2500-2509.

[11] Veldhuizen, R., Nag, K., Orgeig, S., Possmayer, F. "The role of lipids in pulmonary surfactant" (1998) Biochimica et Biophysica Acta 1408 90-108.

[12] Khanorkar, S. V. "Insights in Physiology" (2012) JP Medical Ltd.

[13] Yu, S. H., Possmayer, F. "Lipid compositional analysis of pulmonary surfactant monolayers and monolayer-associated reservoirs" (2003) Journal of Lipid Research 44, 621-629.

[14] Protein data bank current holdings structures:

http://www.rcsb.org/pdb/statistics/holdings.do (accessed: 10/04/2016).

[15] Membrane proteins of known structure:

http://blanco.biomol.uci.edu/mpstruc/ (accessed: 10/04/2016).

[16] Yi, H. Y., Chowdhury, M., Huang, Y. D., Yu, X. Q. "Insect antimicrobial peptides and their applications" (2014) Appl Microbiol Biotechnol. 98(13), 5807-22.

[17] Rajanbabu, V., Chen, J. Y. "Applications of antimicrobial peptides from fish and perspectives for the future" (2011) Peptides 32(2), 415-420.

[18] De Smet, K., Contreras, R. "Human antimicrobial peptides: defensins, cathelicidins and histatins" (2005) Biotechnol Lett. 27(18), 1337-47.

[19] Wang, G., Li, X. and Wang, Z. "APD2: the updated antimicrobial peptide database and its application in peptide design" (2009) Nucleic Acids Res 37, D933-D937.

[20] Nguyen, L. T., Haney, E. F. and Vogel, H. J. "The expanding scope of antimicrobial peptide structures and their modes of action" (2011) Trends Biotechnol 29, 464-472.

[21] Zasloff, M. "Antimicrobial peptides of multicellular organisms" (2002) Nature 415, 389-395.

[22] Brogden, K. A. "Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria?" (2005) Nature Reviews Microbiology 3, 238-250.

[23] Bechinger, B. and Salnikov, E. S. "The membrane interactions of antimicrobial peptides revealed by solid-state NMR spectroscopy" (2012) Chem Phys Lipids 165, 282-301.

[24] Seo, M. D., Won, H. S., Kim, J. H., Mishig-Ochir, T. and Lee, B. J. "Antimicrobial peptides for therapeutic applications: a review." (2012) Molecules 17, 12276-12286.

[25] Yeung, A. T., Gellatly, S. L. and Hancock, R. E. "Multifunctional cationic host defence peptides and their clinical applications" (2011) Cell Mol Life Sci 68, 2161- 2176.

[26] Rossi, L. M., Rangasamy, P., Zhang, J., Qiu, X. Q. and Wu, G. Y. "Research advances in the development of peptide antibiotics" (2008) J Pharm Sci 97, 1060-1070.

[27] Hoskin, D. W. and Ramamoorthy, A. "Studies on anticancer activities of antimicrobial peptides" (2008) Biochim Biophys Acta 1778, 357-375.

[28] Al-Benna, S., Shai, Y., Jacobsen, F. and Steinstraesser, L. "Oncolytic activities of host defense peptides" (2011) Int J Mol Sci 12, 8027-8051.

[29] Possmayer, F. "Physicochemical Aspects of Pulmonary Surfactant " (1998) (Chapter 115). Fetal and Neonatal Physiology. 2nd ed. R.A. Polin, and W.W. Fox. Philadelphia, PA: W.B. Saunders Company. 1259-1275.

[30] Blom, J. A. "Monitoring of Respiration and Circulation" (2004) CRC Press. ISBN 978-0-8493-2083-5.

[31] Ochs, M., Nyengaard, J. R., Jung, A., Knudsen, L., Voigt, M., Wahlers, T., Richter, J., Gundersen, H. J. G. "The number of alveoli in the human lung"

(2004) American journal of respiratory and critical care medicine 169 (1), 120-4.

[32] Wert, S. E., Whitsett, J. A., Nogee, L. M. "Genetic disorders of surfactant dysfunction" (2009) Pediatric and Developmental Pathology 12 (4), 253-74.

[33] Pertutz, M. F. "Structure and function of haemoglobin: I. A tentative atomic model of horse oxyhaemoglobin" (1965) J. Mol. Biol. 13, 646-668.

[34] Pertutz, M. F. "Structure and function of haemoglobin: II. Some relations between polypeptide chain configuration and amino acid sequence" (1965) J. Mol. Biol. 13, 669-678.

[35] Chothia, C., Lesk, A. M. "The relation between the divergence of sequence and structure in proteins" (1986) EMBO J. 5(4), 823-826.

[36] Sadowski, M. I., Jones, D. T. "Benchmarking template selection and model quality assessment for high-resolution comparative modeling" (2007) Proteins: Structure, Function, and Bioinformatics 69 (3), 476-485.

[37] Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L., Schwede, T. "SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information" (2014) Nucleic Acids Research 42 (W1) W252-W258.

[38] Arnold, K., Bordoli, L., Kopp, J., Schwede, T. "The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling" (2006) Bioinformatics 22,195-201.

[39] Kiefer, F., Arnold, K., Kunzli, M., Bordoli, L., Schwede, T. "The SWISS-MODEL Repository and associated resources" (2009) Nucleic Acids Research 37, D387-D392.

[40] Guex, N., Peitsch, M. C., Schwede, T. "Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective" (2009) Electrophoresis, 30(S1), S162-S173.

[41] van der Spoel, D., Lindahl, E., Hess, B., van Buuren, A. R., Apol, E., Meulenhoff, P. J., Tieleman, D. P., Sijbers, A. L. T. M., Feenstra, K. A., van Drunen, R. and Berendsen, H. J. C. "Gromacs User Manual version 4.5.6" www.gromacs.org (2010).

[42] Yates, C., Sharp, S., Jones, J., Topps, D., Coleman, M., Aneja, R., Jaynes, J. and Turner, T. "LHRH-conjugated lytic peptides directly target prostate cancer cells" (2011) Biochem Pharmacol 81, 104-110.

[43] Frenkel, D., Smit, B. "Understanding molecular simulation from algorithms to applications" (1996) Academic Press.

[44] Verlet, L. "Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules" (1967) Phys. Rev. 159, 98.

[45] Verlet, L. "Computer "Experiments" on Classical Fluids. II. Equilibrium Correlation Functions" (1968) Phys. Rev. 165, 201.

[46] Jorgensen, W. L., Maxwell, D. S. and Tirado-Rives, J. "Development and testing of the OPLS all-atom force field on conformational energetics and properties of

organic liquids" (1996) Journal of the American Chemical Society 118, 11225-11236.

[47] Kaminski, G. A., Friesner, R. A., Tirado-Rives, J. and Jorgensen, W. L. "Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides" (2001) J. Phys. Chem. B 105, 6474-6487.

[48] Tieleman, D. P., MacCallum, J. L., Ash, W. L., Kandt, C., Xu, Z. and Monticelli, L. "Membrane protein simulations with a united-atom lipid and all-atom protein model: Lipid-protein interactions, side chain transfer free energies and model proteins" (2006) J Phys Condens Matter 18, S1221-S1234.

[49] Berendsen, H. J. C., Postma, J. P. M., DiNola, A., Haak, J. R. "Molecular dynamics with coupling to an external bath" (1984) J. Chem. Phys. 81, 3684-3690.

[50] Nose, S. A "molecular dynamics method for simulations in the canonical ensemble" (1984) Mol. Phys. 52, 255-268.

[51] Hoover, W. G. "Canonical dynamics: equilibrium phase-space distributions" (1985) Phys. Rev. A 31, 1695-1697.

[52] Berendsen, H. J. C., Postma, J. P. M., DiNola, A., Haak, J. R. "Molecular dynamics with coupling to an external bath" (1984) J. Chem. Phys. 81, 3684-3690.

[53] Parrinello, M., Rahman, A. "Polymorphic transitions in single crystals: A new molecular dynamics method" (1981) J. Appl. Phys. 52, 7182-7190.

[54] Nose, S., Klein, M. L. "Constant pressure molecular dynamics for molecular systems" (1983) Mol. Phys. 50, 1055-1076.

[55] Kutzner, C., Van Der Spoel, D., Fechner, M., Lindahl, E., Schmitt, U. W., De Groot, B. L. and Grubmuller, H. "Speeding up parallel GROMACS on high-latency networks" (2007), J. Comput. Chem., 28: 2075-2084.

[56] Jorgensen, W. L., Maxwell, D. S., Tirado-Rives, J. "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids" (1996) J. Am. Chem. Soc. 118, 11225-11236.

[57] Ewald, P. P. "Die berechnung optischer und elektrostatischer gitterpotentiale" (1921) Ann. Physik, 64, 253. Translation: "The computation of optical and electrostatic lattice potentials".

[58] Darden, T., York, D., Pedersen, L. "Particle mesh Ewald: An N log(N) method for Ewald sums in large systems" (1993) J. Chem. Phys. 98, 10089.

[59] Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., de Vries, A. H. "The MARTINI forcefield: coarse grained model for biomolecular simulations" (2007) J. Phys. Chem. B 111, 7812-7824.

[60] Sindhikara, D., Meng, Y., Roitberg A. E. "Exchange frequency in replica exchange molecular dynamics" (2008) The Journal of Chemical Physics 128, 024103.

[61] Introduction to REMD using GROMACS:

http://www.gromacs.org/Documentation/How-tos/REMD
(accessed: 18/09/2015).

[62] Bechinger, B., Sizun, C. "Alignment and structural analysis of membrane polypeptides by ^{15}N and ^{31}P solid-state NMR spectroscopy" (2003) Concepts in Magnetic Resonance Part A, 18A(2), 130-145.

[63] Duer, M. J. "Introduction to solid state NMR spectroscopy" (2004) Blackwell publishing.

[64] Aisenbrey, C., Sudheendra, U. S., Ridley, H., Bertani, P., Marquette, A., Nedelkina, S., Lakey, J. H. and Bechinger, B. "Helix orientations in membrane-associated Bcl-X(L) determined by ^{15}N solid-state NMR spectroscopy" (2007) Eur Biophys J. 37, 71-80.

[65] Vosegaard, T., Kamihira-Ishijima, M., Watts, A. and Nielsen, N. C. "Helix Conformations in 7TM Membrane Proteins Determined Using Oriented-Sample Solid-State NMR with Multiple Residue-Specific ^{15}N Labeling" (2008) Biophysical Journal 94, 241-250.

[66] Frishman, D., Argos, P. "Knowledge-Based Protein Secondary Structure Assignment" (1995) Proteins: Structure, Function, and Genetics 23, 566-579.

[67] Touw, W. G., Baakman, C., Black, J., te Beek, T. A., Krieger, E., Joosten, R. P., Vriend, G. "A series of PDB related databases for everyday needs" (2015) Nucleic Acids Research 43 (Database issue), D364-D368.

[68] Kabsch, W., Sander, C. "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features" (1983) Biopolymers 22, 2577-2637.

Chapter 2

Molecular dynamics simulations of histidine-containing cod antimicrobial peptide paralogs in self-assembled bilayers[1]

2.1 Introduction

Antimicrobial peptides (AMPs) play an important role in the immune systems of a wide variety of organisms, from humans to fish to insects. AMPs are generally amphipathic and cationic in nature [1], and consequently have a propensity to interact with lipid bilayers. Much of the research into the mechanism of AMPs has focused

[1]This chapter is published as an article in Biophysica et Biochimica Acta (BBA)-Biomembranes (2014).

on their interaction with membranes, either as their direct mechanism of killing, or as a means of getting inside the cell to disrupt intracellular targets [2–5]. Besides their natural roles in innate immunity, AMPs are also being investigated as potential therapeutics for conditions such as drug resistant infection [6–8] and cancer [9–11].

Many AMPs exhibit a degree of specificity and can kill pathogens at concentrations that do not harm host cells. At least some of this specificity is believed to derive from their cationic nature which provides for stronger interactions with, for example, bacterial or cancer cell membranes, which are generally more anionic in character than mammalian host cells [12, 13]. Nonetheless, one major barrier to using AMPs as drugs is that at high concentrations, they can kill not just the target cells but the healthy host cells as well [14, 15].

One way of controlling the specificity and activity of AMPs is via controlling their charge. This is particularly applicable to peptides that contain the amino acid histidine, which is generally uncharged at neutral pH but tends to become positively charged at mildly acidic pH. This pH sensitivity of histidine-containing AMPs can provide a "pH switch" to activate them in lower pH environments [16–19]. Acidic pH activation of AMPs may play a role in the endogenous functioning of AMPs; for example, acidic pH is important in skin immune defence [16–20]. Moreover, there are intriguing possibilities to employ pH-activated AMPs in exogenous applications, such as the treatment of cancer [9, 10] as the environment around tumors is usually acidic [17, 21]. Thus, histidine-containing AMPs are excellent candidates as therapeutics because they are likely to be much more active in the vicinity of the tumor than elsewhere in the body. Indeed, replacement of arginines and lysines with histidine in the AMP K6L6 was shown to make the peptide more specific; systemic injection of

the modified peptide inhibited tumor growth in mice with reduced systemic toxicity compared to the parent peptide [17].

At least three different mechanisms have been suggested for the altered activity of histidine-containing AMPs that is observed when histidines become more positively charged at lower pH. 1) Membrane binding: Kacprzyk et al. [16] replaced the lysine and arginine in synthetic AMP sequences with histidines to produce peptides that were only active under acidic conditions. The differences in activity corresponded well to observed differences in membrane binding. 2) Membrane penetration: Kharadia et al. [22] also replaced the lysine and arginine residues in lytic peptides with histidines. They found that the novel peptides were much more selective for bacteria over host tissue cells and attributed the increase in activity at lower pH not to changes in membrane binding, but to increased membrane penetration. 3) Changing peptide structure: This mechanism for pH-dependent activity in histidine-containing AMPs was suggested by detailed studies of the synthetic peptide LAH4, for which pH alterations induce changes in the peptide structure. At neutral pH a long helix is formed, which associates with the membrane in a planar manner. At slightly acidic pH, a hinge disrupts the contiguous helix structure, possibly due to electrostatic repulsion between adjacent histidines [23].

We were interested in exploring, at an atomic level, the role of histidines in AMP-membrane interactions. Of particular interest were histidines in natural sequences as opposed to the synthetic sequences that have been the subjects of most mechanistic studies of histidine containing AMPs so far. Additionally, we wanted to probe the potential significance of histidines that appear in sequential pairs in the sequence, as opposed to histidines flanked solely by non-histidine residues. The peptides employed

in the study, GAD-1 and GAD-2, are derived from codfish sequence libraries [3–6]. They are paralogs, i.e., related genes found in the same organism, and their comparison has potential to illuminate the role of histidine pairs in evolutionarily tuned structure-function relationships. GAD-1 and GAD-2 are members of the piscidin family of helical AMPs. Piscidins have been subjects of several structure-function studies, although not studies that address pH-dependent behavior and mechanisms [28–37]. GAD-1 (FIHHIIGWISHGVRAIHRAIH-NH2) has 5 histidines, two of which appear in a pair and three of which appear singly. GAD-2 (FLHHIVGLIHHGLSLFGDR-NH2) has 4 histidines, which appear in two sets of histidine pairs. We studied these peptides in histidine charged forms, denoted as GAD-1p and GAD-2p, as well as in the histidine neutral form (GAD-1, GAD-2). All-atom molecular dynamics simulations with these 4 peptides along with POPC lipids were performed in order to reveal atomistic details of their lipid interactions. As detailed in the Methods and Discussion sections, with our system setup, pores form even in the absence of peptide, and thus our study does not provide a kinetic picture of how the peptides bind to membranes and induce pore formation, but rather provides details of the peptide/lipid interactions. One of our key findings was that histidine pairs are more likely to be found closely associated with the pore than in the more ordered, planer region of the lipid bilayers.

2.2 Methods

Our approach followed the method of Salgado et al. [7] who start unassembled lipid molecules in random positions with a single peptide among the lipid molecules. This method allows the system to freely assemble into a peptide-bilayer complex, thus

avoiding any artifacts that might result from, for example, introducing the peptide into a pre-assembled bilayer after removing lipids "by hand". The main difference in our implementation of this approach is that, rather than employing a script to randomize the position of the lipids, we randomized the system by simulating at high temperature, 1400 K. This modification provided ease of implementation, in particular by allowing for an initial volume not too much larger than the final one, while avoiding truly unphysical interactions, e.g. overlapping molecules.

Salgado et al. [7] found that the method produced expected outcomes in terms of the location and orientation of the peptide with respect to the membrane, i.e., a hydrophobic peptide ended up in a transmembrane configuration, while a more amphipathic peptide ended up on and parallel to the membrane surface. This lends confidence that the assembled structures reflect low free energy states of the equilibrium system. By contrast to these two simpler cases, imperfectly amphipathic peptides, such as the ones we are studying,may possess many different configurations of similar free energy. To address this, for each system composition,we performed four independent simulations of the self-assembly process. While computational resources limited us to this small number, it did provide sufficient sampling for at least a semi-quantitative characterization of the differences between the paralogs in their charged and uncharged forms. Additionally, simulations were carried out in the absence of the peptide to control for the effect of the peptide on bilayer formation.

The systems consisted of 128 POPC lipid molecules (6656 atoms), a single peptide ($\sim$ 350 atoms) and approximately and no less than 7360 water molecules (roughly 37,000 atoms in total) (**Table 2.1**). This number of lipid molecules provided a large enough bilayer to accommodate the peptide in the presence of a pore. The

System	Number of atoms	Time (ns)	Class type	Bilayer width (nm)	Pore diameter (P) (nm)	Pore diameter (W) (nm)
POPC-A	36,368	400.00	–	4.24±0.09	0.69±0.16	1.08±0.13
POPC-B	36,368	400.00	–	4.22±0.09	0.66±0.16	1.05±0.13
POPC-C	36,368	400.00	–	4.27±0.09	0.52±0.18	0.98±0.14
POPC-D	36,368	400.00	–	4.26±0.08	0.71±0.18	1.09±0.13
GAD-1-A	36,728	376.00	1	4.25±0.09	0.47±0.17	0.98±0.18
GAD-1-B	36,728	370.40	1	4.15±0.10	0.53±0.18	1.03±0.14
GAD-1-C	36,728	348.00	5	4.31±0.09	0.59±0.16	1.00±0.12
GAD-1-D	36,728	344.00	2	4.21±0.08	0.37±0.15	0.92±0.23
GAD-1p-A	36,498	344.00	1	4.17±0.08	0.62±0.21	1.18±0.21
GAD-1p-B	36,498	320.40	1	4.22±0.08	0.82±0.14	1.14±0.11
GAD-1p-C	36,498	328.00	4	4.28±0.09	0.81±0.16	1.13±0.11
GAD-1p-D	36,498	304.00	1	4.24±0.09	0.57±0.17	1.01±0.26
GAD-2-A	36,576	336.00	3	4.01±0.08	–	–
GAD-2-B	36,576	332.40	2	4.25±0.09	0.77±0.21	1.18±0.10
GAD-2-C	36,576	335.00	2	4.25±0.09	0.52±0.20	1.15±0.13
GAD-2-D	36,576	324.00	2	4.19±0.08	0.40±0.20	1.04±0.17
GAD-2p-A	36,624	323.60	3	4.22±0.09	0.74±0.18	1.10±0.13
GAD-2p-B	36,624	332.00	2	4.26±0.09	0.79±0.12	1.21±0.13
GAD-2p-C	36,624	348.00	4	4.24±0.10	0.64±0.16	1.01±0.13
GAD-2p-D	36,624	328.00	1	4.21±0.09	0.58±0.16	1.06±0.15

Table 2.1: Parameters and measurement summary for each simulation. The calculations for bilayer width, pore diameter (P) and pore diameter (W) were done for last 100 ns of simulations. A, B, C and D represent individual runs of the same system composition.

amount of water was chosen to allow for sufficient space to prevent periodic boundary conditions from permitting the peptide to interact unphysically with both leaflets through water. The number of water molecules employed is somewhat larger than the number employed in other studies [7, 39, 40].

The initial alpha-helical peptide structures were generated using Swiss PDB Viewer (SPV) [41–44]. The C-terminus was amidated with an NH_2 group to be in line with

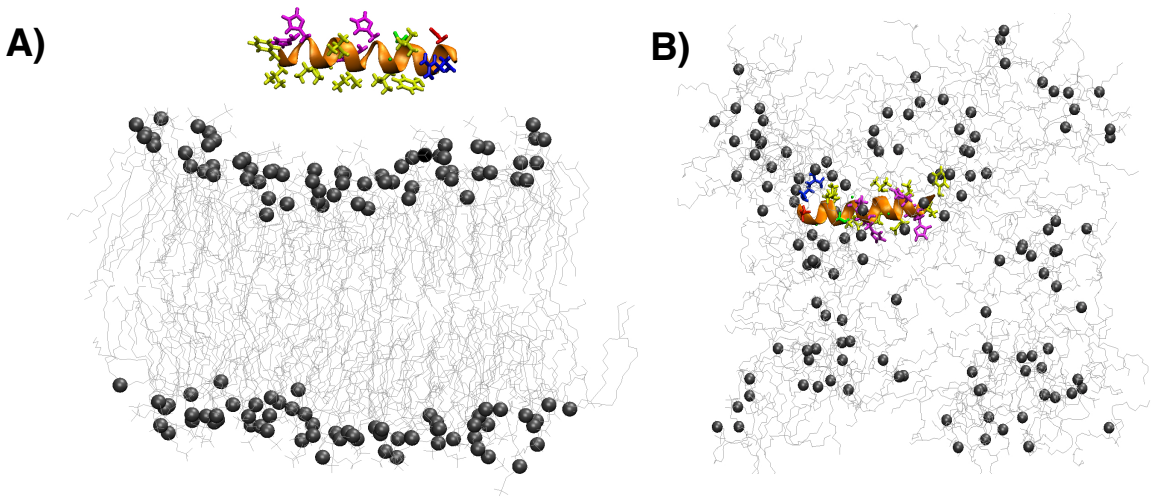

Figure 2.1: System set up. The peptide was initially configured as a canonical helix and placed in the simulation box with an assembled POPC lipid bilayer (A). Keeping the peptide position restrained, the molecule positions were randomized by heating the system to 1400 K for 2 ns (B). The peptide backbone is shown as an orange ribbon, the histidine side chains in purple, hydrophobic side chains in yellow, polar uncharged side chains in green, positively charged side chains in blue, and negatively charged side chains in red. The gray spheres represent the lipid headgroup phosphorus atoms and the silver lines, the lipid acyl-chains.

ongoing experimental work. The peptide was placed in a cubic simulation box of side length 8 nm along with the 128 POPC molecules, which were initially arranged in a bilayer configuration taken from [45] (**Figure 2.1-A**). Sufficient Cl^- counterions were added to ensure overall charge neutrality: 3 ions for GAD-1, 8 for GAD-1p, 1 for GAD-2 and 5 for GAD-2p. For histidines in GAD-1p and GAD-2p, the protonated form of histidine was used.

GROMACS version 4.5 was used for the simulations [46]. We employed a version of the all atom Optimized Potential for Liquid Simulations (OPLS-AA) force field [15,

16], adapted for POPC lipid molecule properties [49].

As the first step in generating randomized starting configurations for the self-assembly process, we carried out a simulation of the system comprising the peptide, lipid bilayer and counter ions in the canonical ensemble at T=1400 K for 2 ns, still within a cubic box of side length 8 nm, while restraining the position of all the peptide atoms to preserve its helical structure. We employed the modified Berendsen thermostat (v-rescale in GROMACS). Under these conditions, the lipid bilayer immediately disassembles, equilibrating rapidly to a highly mobile fluid of lipid molecules (**Figure 2.1-B**). In this regime, the root mean square displacement of a lipid molecule over 1 ns is approximately 10 nm.

Next, we added approximately 7500 TIP4P water molecules [50] to the simulation box and, after an energy minimization, continued running at 1400 K with the peptide still restrained. The root mean square displacement of lipid molecules over 2 ns was approximately 8.5 nm. During this run, we harvested four configurations, one every 2 ns, that served as independent starting configurations for separate realizations of the self-assembly process,which we label as A, B, C and D. This procedure for obtaining four independent configurations was carried out for each variant of the peptide (GAD-1, GAD-1p, GAD-2 and GAD-2p) as well as for a system without a peptide, which acts as a control. In total, there were starting points for 20 simulations of the self-assembly process.

Each self-assembly simulation began with a brief simulation of 100 ps under conditions of constant temperature and constant pressure, with the peptide unrestrained. The temperature was held constant at 310 K with the Nose-Hoover algorithm and a time constant of 0.1 ps. An isotropic pressure of 1 bar was maintained with the

Parrinello-Rahman algorithm employing a time constant of 5 ps and compressibility of 4.5×10^{-5} bar^{-1}. This short simulation was sufficiently long to bring the density into a steady state (it does not continue to evolve in subsequent steps), as well as to allow the large fluctuations expected after a large drop in temperature to subside. This run was short enough so that no significant change in the helicity of the peptide occurred and the lipid molecules did not show any appreciable progress towards bilayer assembly. The final box size was roughly 7.25 nm on each side.

At this point we began the "production" run (**Figure 2.2**), using a time step of 2 fs and a radial cutoff of 1.2 nm for the real space force calculations. We used the particle mesh Ewald method for Columbic interactions with a Fourier spacing of 0.15 nm and interpolation order of four (cubic). The temperature and pressure controls were as for the previous step, with the exception that we use anisotropic pressure scaling in order to avoid enforcing a bilayer normal direction (as opposed to semi-isotropic). The average run time for these production simulations was approximately 350 ns (**Table 2.1**). The beginning of these runs marks our time origin (t= 0).

We note that in some of our runs, anisotropic scaling allowed for one of the dimensions of the box to shrink below the potential cutoff, effectively ending the simulation. In retrospect, we could have used anisotropic rescaling until the bilayer had formed, and then switched to semi-isotropic scaling to avoid this problem. In reality, we generated additional starting configurations so as to have four realizations of each variant and to maintain a uniform protocol across runs.

In the "Results" and "Discussion" sections, the formation and presence of pore in lipid bilayer generated with self-assembly method will be explained and discussed. To investigate the formation of pore in our lipid bilayers, in addition to POPC-A,

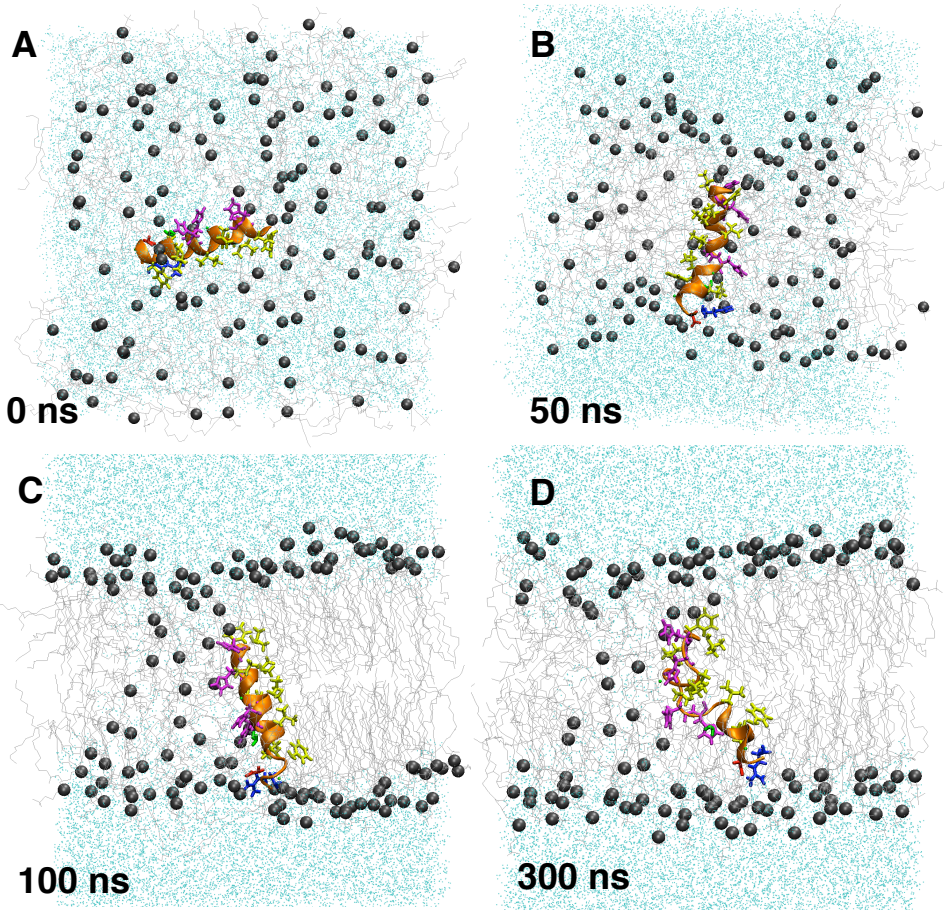

Figure 2.2: Representative snapshots of the molecular dynamics simulation of GAD-2 peptide realization set C in POPC. (A) At the start of the production run (0 ns) with randomized lipid positions. (B) After 50 ns, a more ordered, bilayer-like configuration is observed. (C) The peptide is positioned inside a bilayer (100 ns). Although the bilayer appears ordered with sharply defined lipid head group positions, a pore is present. (D) After 300 ns, the bilayer still has a pore and the peptide is found inside the bilayer, proximal to the pore. The peptide backbone is shown as an orange ribbon, the histidine sidechains in purple, hydrophobic sidechains in yellow, polar uncharged sidechains in green, positively charged sidechains in blue, and negatively charged sidechains in red. The gray spheres represent the lipid headgroup phosphorus atoms, the silver lines represent the lipid acyl-chains, and water is shown as cyan dots.

POPC-B, POPC-C and POPC-D systems, we set up four other sets of simulations, named Test-1, Test-2, Test-3 and Test-4 at T=323 K to help us study the effect of pressure coupling, temperature change and water model choice, on generation and stability of pore in lipid bilayer during the course of simulation. The temperature is set to 323 K for our simulations to be comparable with previously studied systems [7, 62, 63]. In Test-1 and Test-2 simulations, we used the final structure of one of our 4 peptide-free control simulations as our initial structure and run our simulations. In the Test-1 system, we continue our simulation with the same parameters as our control simulations, except T=323 K. This simulation ran for 73 ns until one of the dimensions of simulation box shrank below the potential cutoff. In the Test-2 system, in addition to setting T=323 K, we change the pressure coupling to semiisotropic. For Test-3 and Test-4 systems, we run our simulations from the beginning, with the same simulation parameters, that we change the water model from TIP4P to SPC water molecule [51]. In Test-3, we added $\sim$ 11000 SPC water molecules, while in the Test-4, we used 6000 SPC water molecules. Test-3 ran for 96 ns and Test-4 ran for 49 ns.

To characterize the evolution and structure of all our systems, we tracked the potential energy of the system, lipid acyl-chain order parameter (as defined in Salgado et al. [7]), the width of lipid bilayer, pore size, secondary structure of the peptide, mean square displacement of particles and the number of water molecules around different peptide residues. To find the mean value of order parameter of acyl-chains as a function of time, for each frame we calculated the order parameter of each of the carbons in the unsaturated acyl-chain and averaged over all of these carbons. The utilities included with GROMACS, such as g_energy, g_density, g_order, g_msd

and g_rdf, as well as the Define Secondary Structure of Proteins (DSSP) [52, 53] plugin of the Visual Molecular Dynamics (VMD) software package [54, 55] were used to obtain these quantities.

To determine the pore size, we first needed to determine the location of the pore axis. To do this, we found all interior P atoms, i.e., those residing within 1.5 nm of the midway plane between bilayer leaflets. (We defined the location of a leaflet along the bilayer normal as a peak in the P density profile). On average, for a given simulation frame, there are six such interior P atoms. The pore axis, which by definition is parallel to the bilayer normal, passes through the center of mass of the interior P atoms and was defined on a frame-by-frame basis. The radius of the pore, by one definition, was the average perpendicular distance of the interior P atoms to the axis. Alternatively, we defined the pore size by considering all distances between the interior water molecules (defined in the same way as interior P atoms), and finding the largest perpendicular distance. Since there was the occasional water molecule that diffused deep into the bilayer, in calculating the average pore size, we discarded the largest 5% of sizes. We determined this cutoff by looking at the distribution of pore sizes.

To calculate the percentage of time each peptide residue takes on a helical structure we used VMD software (version 1.9.1). We found the secondary structure of each residue frame-by-frame in VMD, and then used this to calculate the percent helicity per residue during the last 100 ns of the simulation (except for GAD-1p-A, where the last 70 ns was used). Both alpha-helical and 3-10 helical structures were used to generate the percent helical structure.

2.3 Results

2.3.1 Peptide-free Systems

In peptide-free systems, we ran two types of simulations in which the system does not contain peptide. For the first type, we used the main protocol to run our simulations (POPC-A, POPC-B, POPC-C and POPC-D) and for the second type, we studied the effect of temperature, pressure scaling and water model on the formation and healing of the pore in POPC lipid bilayer.

2.3.1.1 Main Protocol Simulations

We first investigated the behavior of the POPC/water systems in the absence of peptide (POPC-A, POPC-B, POPC-C, POPC-D in **Table 2.1**). After starting with the lipids in randomized positions, all four simulations reach apparent equilibrium after $\sim$ 100 ns, as judged by multiple parameters, including potential energy and the mean value of order parameter of the lipid chains (**Figure 2.3**). From 100 ns to the end of the simulations at $\sim$ 400 ns, all 4 systems display a bilayer with a single pore. The pore consists of a contiguous water-containing hole through both leaflets of the bilayer that is lined by phospholipid headgroups. The area per headgroup of the lipids lining the pores is larger than the area per headgroup for lipids found in the planar region of the bilayer.

The bilayer width was found to be between 4.2 and 4.3 nm (**Table 2.1**) as defined by the distance between the phosphorous atoms in the lipid head groups. The sizes of the pores varied from simulation to simulation, but were in the range of 1.0 to 1.1 nm when measured using water, and in the range of 0.5 nm to 0.8 nm when measured

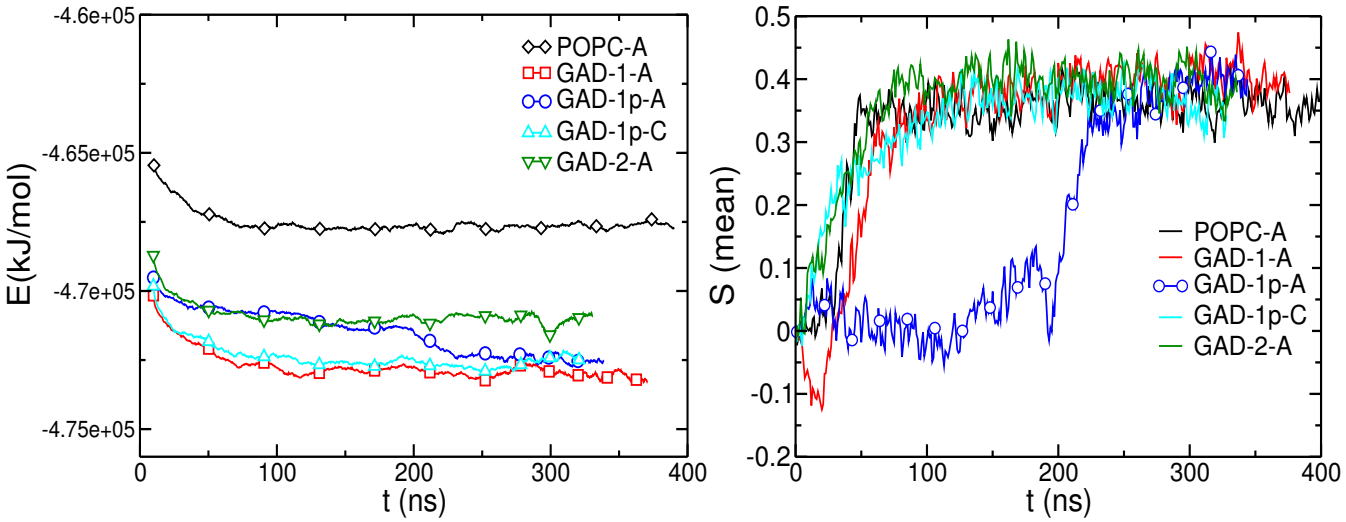

Figure 2.3: Time evolution of potential energy (left) of the system and mean value of the order parameter of the acyl-chain (right) during representative simulations. In the top panel, shown are running averages over 50 data points, with 400 ps between data points.

using phosphorous (**Table 2.1**). The difference in pore size for the two measurements illustrates that water penetrates well into the lipids of the pore. As we shall discuss, this more aqueous environment plays an important role in understanding the structure and positioning of the peptide near the pore.

2.3.1.2 Pore-formation-test Simulations

To analyze the behaviour of pore with different temperatures, pressure scaling and water model, we studied 4 pore-formation-test simulations (Test-1, Test-2, Test-3 and Test-4 in **Table 2.2**) at T=323 K, in addition to the control simulations we run in **Table 2.1** (POPC-A, POPC-B, POPC-C and POPC-D) at T=310 K. In Test-1 and Test-2, the system keeps the initial pore inside the bilayer during simulation. The bilayer width is smaller compared to the control simulations run at 310 K. In Test-2,

System	Number of atoms	Water	Pressure	Bilayer width (nm)	Bilayer(ns)	Healing (ns)	Duration (ns)
Test-1	36,368	TIP4P	anisotropic	4.14±0.10	–	–	73
Test-2	36,368	TIP4P	semiisotropic	4.08±0.09	–	–	25.2
Test-3	40,685	SPC	anisotropic	4.21±0.10	20	82	104
Test-4	24,656	SCP	anisotropic	4.13±0.16	32	48	49

Table 2.2: Parameters and measurement summary for pore-formation-test simulations.

changing the pressure scaling to semiisotropic did not help the pore to heal. In Test-3 and Test-4, we replaced the TIP4P water molecules with SPC model water. We can see that in these simulations the bilayer forms with a pore inside it earlier compared to the systems with TIP4P water models and, strikingly, the pore heals during the simulation. The smaller number of water molecules in the Test-4 system, could be the reason that the pore heals faster compared with the Test-3 simulation. Thus, the choice of water model has a strong influence on the longevity of the pore.

2.3.2 Systems with Peptides

Next, we added peptide to our systems. We studied four peptides, GAD-1, GAD-1p, GAD-2, and GAD-2p, where "p" denotes the form of the peptide with histidines positively charged. Each peptide was subjected to four independent simulations A, B, C and D (**Table 2.1**). Initially, during randomization, the peptides started in a canonical helical structure but were subsequently simulated at ambient conditions without restraints (**Figure 2.2**).

Similar to the lipid-only simulations, 15 out of 16 of the peptide-lipid simulations reach apparent equilibrium after $\sim$ 100 ns. There was one outlier; the GAD-1p-A system took $\sim$ 230 ns to generate the bilayer (**Figure 2.3**). The GAD-1p-A system

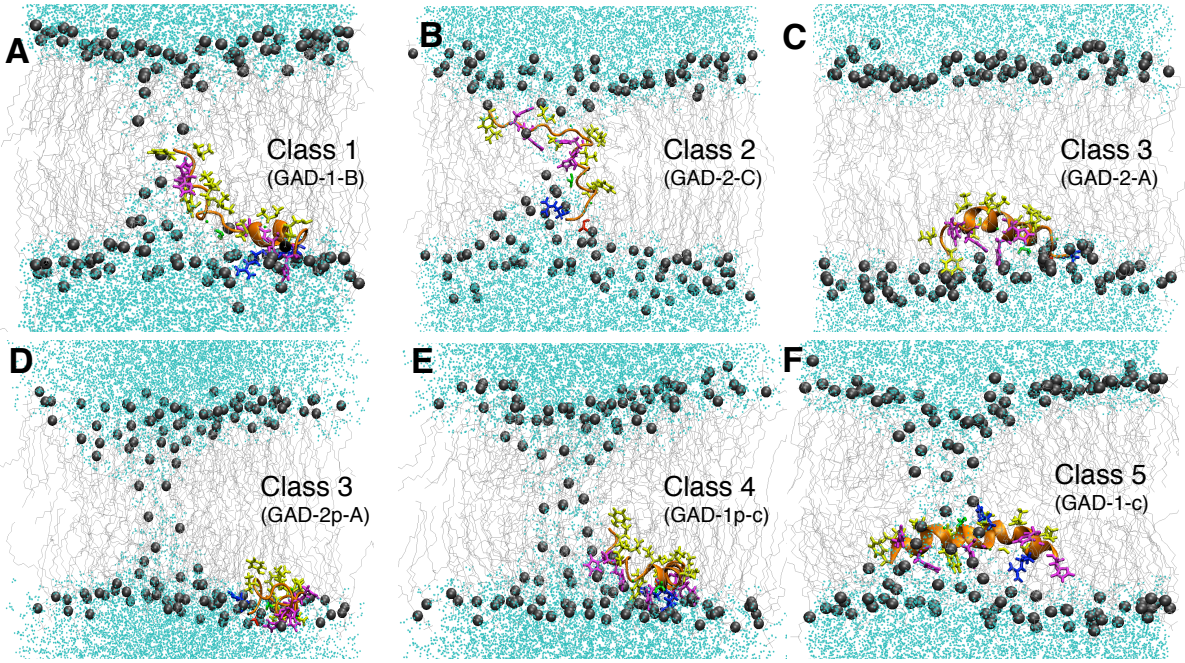

Figure 2.4: (A-F): Selected snapshots from near the end (∼350-400 ns) of the simulations. The peptides exhibit a variety of final structures and topologies within the bilayer, but the systems can be grouped into five types of configurations as described in the text and labeled on the panels. The peptide backbone is shown as an orange ribbon, the histidine sidechains in purple, hydrophobic sidechains in yellow, polar uncharged sidechains in green, positively charged sidechains in blue, and negatively charged sidechains in red. The gray spheres represent the lipid headgroup phosphorus atoms, and the silver lines, the lipid acyl-chains. Water molecules are in cyan.

initially formed with two separate micelles, which eventually coalesced into one micelle and then formed the bilayer. All but one of the peptide-containing systems formed pores (**Figure 2.4**). The one exception was GAD-2-A, the only simulation with a perfect bilayer without a pore (**Figure 2.4-C**).

As for the systems without peptide, the sizes of the pores varied from simulation to simulation, but were in the range of 1.0 to 1.2 nm when measured using water and in the range of 0.5 nm to 0.8 nm when measured using phosphorous (**Table**

2.1). These are similar to pore sizes in the systems without peptide. The bilayer width, defined by the distance between the phosphorous atoms in the head groups, is between 4.0 and 4.3 nm, i.e. also in the same range as for systems without peptide. Thus, the peptide does not appear to have any large effect on the pore size or bilayer width.

As detailed below, in most cases (13 out of 15), the peptide preferentially interacts with the pore, rather than with the planar region of the bilayer. One reason for this preference may be the decreased density of lipid head groups in the pore ($\sim$0.8 nm^{-2}) as compared to the planar region ($\sim$1.5 nm^{-2}). The reduced head group density in the pore may allow the peptide enough space to position in such a way so as to interact optimally with both the hydrophobic region of the lipids and the hydrophilic headgroup/water region.

2.3.3 Peptide Structure and Residue Hydrophobic/Hydrophilic Partitioning

During the self-assembly process, there were no restraints on the peptide, allowing its structure to evolve freely as the bilayer formed. Following apparent equilibrium at $\sim$ 100 ns, the structure of each peptide continued to evolve, but did not change significantly in the last 100 ns. When compared to each other, the final structures of the peptides exhibit substantial heterogeneity i.e., even for different independent simulations of the same peptide, there are different degrees of overall helicity and variations in the regions that are helical (**Figure 2.5**). Despite the structural heterogeneity, the partitioning of each residue into either a hydrophobic or hydrophilic environment is

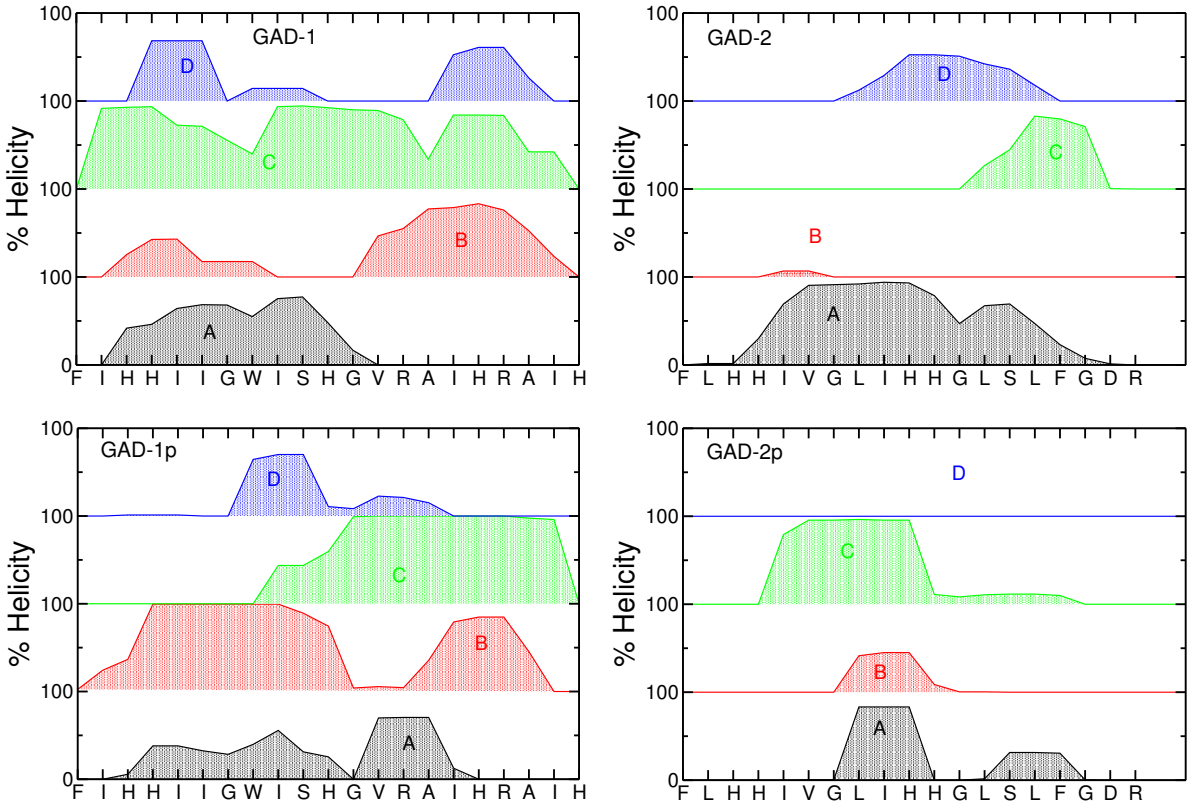

Figure 2.5: Percentage of helical structuring of GAD peptides during last 100 ns of simulation (except GAD-1p-A which is for the last 70 ns). A, B, C and D represent individual runs of the same system composition.

very similar. Specifically, the number of water molecules within a set distance of the center of mass of each residue (**Figure 2.6**) is quite consistent across the 4 different independent simulations of each system. The observed structural heterogeneity is expected and is consistent with both experiments on membrane-active peptides [56] and simulations of AMPs [39, 57]. There is no apparent difference between how the peptides with protonated histidines versus those with neutral histidines behave in terms of helicity and proximity to water.

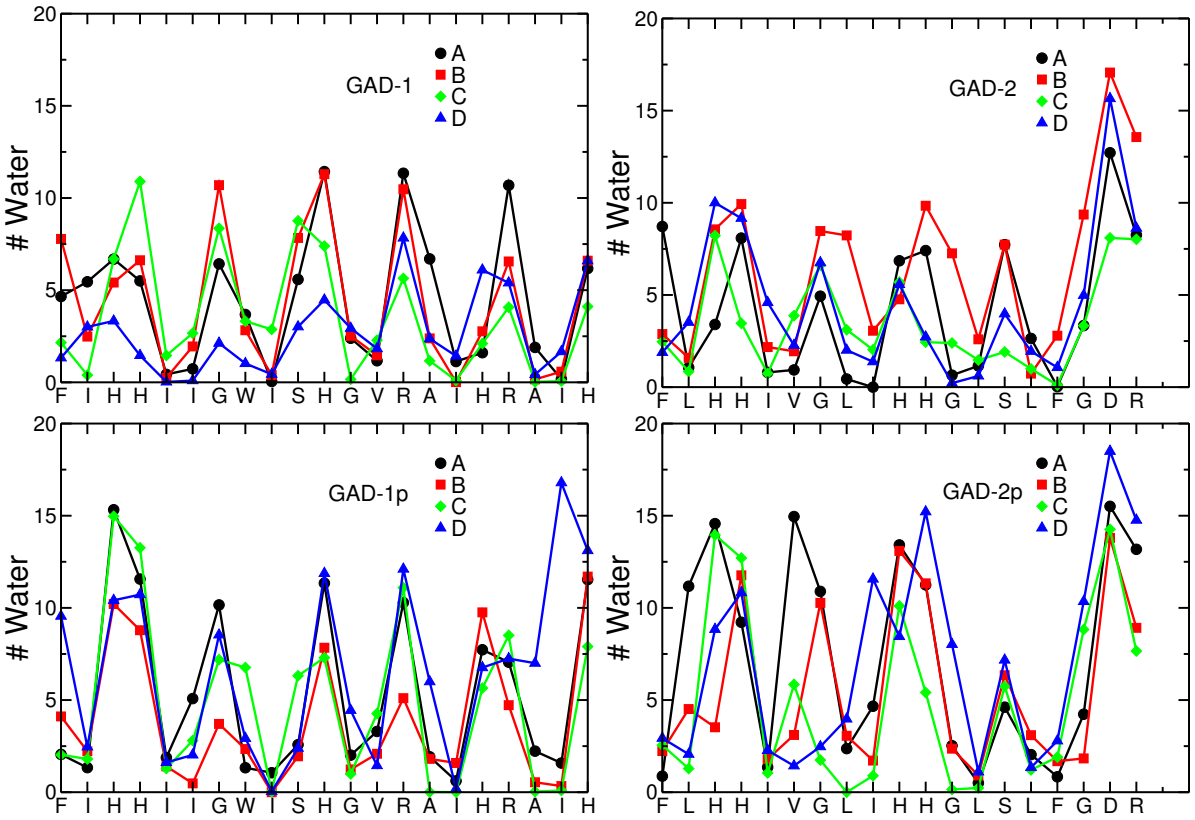

Figure 2.6: Number of water molecules within r=0.65 nm of the center of mass of each GAD residue in the last 100 ns of simulation (except GAD-1p-A which is for the last 70 ns). A, B, C and D represent individual runs of the same system composition.

2.3.4 Peptide-pore Interactions

Of the 15 out of 16 peptide simulations that form a pore, 14 of these show the peptide interacting closely with the pore. This is evident from **Figure 2.7**, where all but one of the peptides (GAD-2p-A) has at least one residue with its center of mass within 0.7 nm of the pore (represented by the dotted lines).

The systems can be classified into five different types, depending on the mode of interaction between the peptide and the pore in the last 100 ns of each simulation. In the first and most common type (6/16 systems), there are extensive interactions

between the pore and one of the terminal halves of the peptide, while the other half of the peptide remains relatively distal from the pore and in the planar region of the bilayer (**Figure 2.4-A**). The region of the peptide that interacts closely with the pore takes on an oblique angle with respect to the plane of the membrane, while the non-interacting peptide half positions more nearly parallel to the membrane surface. The second most common type of system (5/16 systems) is exemplified in **Figure 2.4-B**, where the entire peptide is proximal to the pore and makes extensive interactions with lipids from both leaflets of the bilayer. In the third type of system (2/16), the peptide interacts with the planar region of one of the bilayer leaflets without interacting with a pore, e.g. **Figure 2.4-C** and **Figure 2.4-D**. The fourth class (2/16) consists of systems (GAD-1p-C and GAD-2p-C) in which just one residue in the peptide (H3) interacts with the rim of the pore and the rest of the peptide is not in contact with the pore, but positions within the nearest leaflet, **Figure 2.4-E**. The remaining simulation, GAD-1-C, does not fit within any of these 4 schemes and thus is the sole occupant of the Class 5 type of system. Here, the peptide embeds deeply in the bilayer, but with a position nearly parallel to the bilayer surface.

It appears that, at least on the timescale of these simulations, there is no inter-conversion between class types during the simulation for class 1 and class 2 systems. On the other hand, there were two observed cases where systems started out as class 3 type - i.e. peptide not interacting with the pore - converting to a class 4 type - peptide interacting with the rim of the pore - during the course of the simulation.

The four different peptides, GAD-1, GAD-1p, GAD-2, and GAD-2p show different propensities to form these five types of systems. GAD-1 and GAD-1p both showed a strong tendency to display class 1 configurations, where one end of the peptide

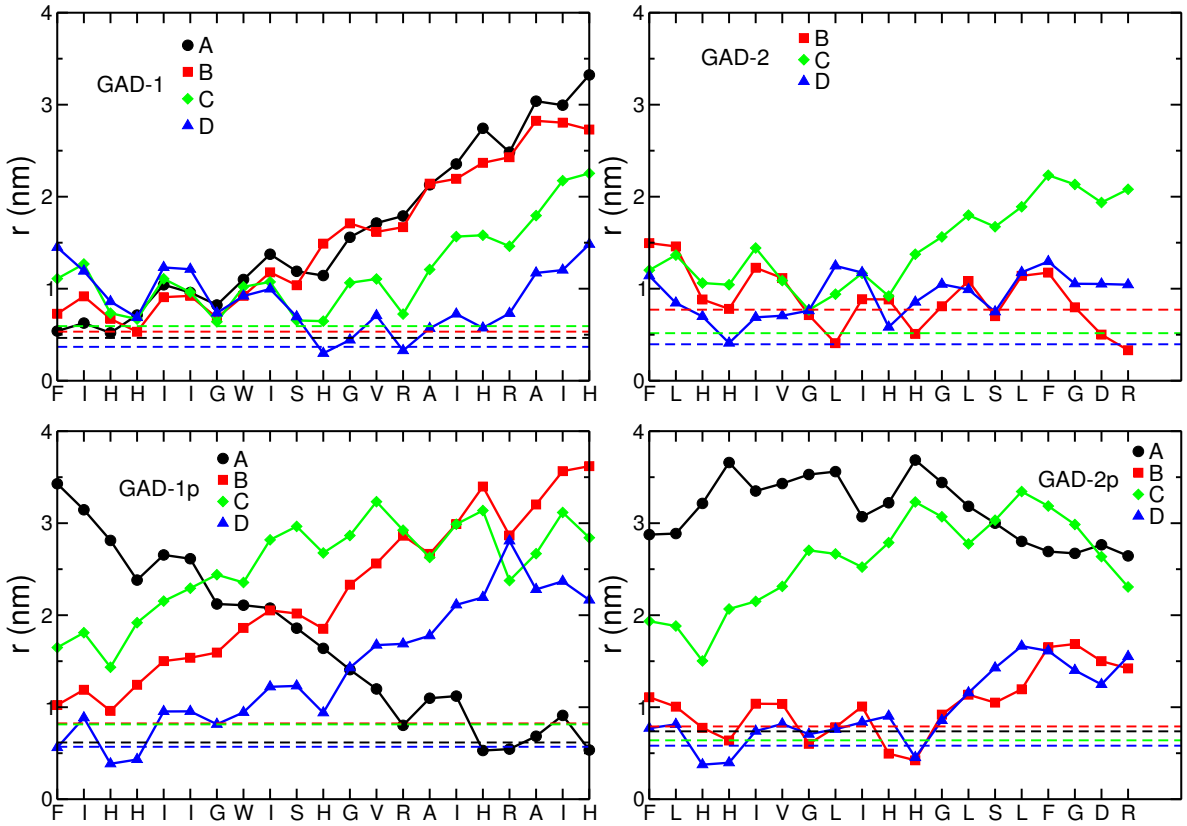

Figure 2.7: The distance of the center of mass of each residue from the central axis of the pore (lines and symbols) and the radius of pore (dotted lines), as defined by lipid phosphorous position during last 100 ns (except GAD-1p-A which is for last 70 ns) of simulation. A, B, C and D represent individual runs of the same system composition.

interacts with the pore, and the other end interacts with the planar region of one bilayer leaflet (**Figure 2.4-A**). One of the exceptions to this trend was also interesting; GAD-1-C is a unique configuration among the 16 peptide-containing systems (**Figure 2.4-F**) with the peptide located in proximity to the pore, deeply embedded in the bilayer with a position parallel to the bilayer normal. Intriguingly, this system has the highest helical structure among all simulations.

In contrast to GAD-1 and GAD-1p, the GAD-2 systems are in class two config-

urations,with the exception of GAD-2-A. In the GAD-2-A system there is no pore. Notably, GAD-2-A also has the most helical structure among GAD-2 simulations. As explored in the "Discussion" section, it may well be significant that the two most helical peptides, GAD-1-C and GAD-2-A, also display the most parallel orientation with respect to the bilayer surface. The GAD-2p sets of simulations are the most heterogeneous in terms of the observed topology and we observe four different types of systems (**Table 2.1**).

2.3.5 Role of Histidine Pairs in AMP-Pore Interactions

One notable observation is that the N-terminal half of all 4 peptides has a greater tendency to interact with the pore than the C-terminal half (**Figure 2.7**). Again the exceptional case is GAD-1p-A, in which the C-terminal half interacts with the pore. A potential explanation for the strong preference for the N-terminal half to interact with the pore is the pair of histidines, (H3 and H4) that are located near the N-terminal half of both peptides. Consistent with this explanation, the 2nd pair of histidines (H10 and H11) present only in GAD-2, are also found consistently in close proximity to the pore (**Figure 2.7**). It appears that there is no need for the histidine pair to be charged in order to interact closely with the POPC pore; even uncharged histidine pairs exhibit this behavior.

To assess the importance of the pairing of the histidines to pore interactions, we can compare the behavior of the H3-H4 pair to the only other charged pair of amino acids in the peptides, the H17 and R18 present in GAD-1p. The H17-R18 pair has a markedly lower tendency to interact with the pore compared to either the H3-H4

motif present in both peptides or the H10-H11 motif present in the GAD-2 peptides (**Figure 2.7**). Specifically, 7 out of 8 of the GAD-1/1p systems show much closer interactions between the pore and the H3-H4-containing N-terminal region of the peptide compared to the H17-R18-containing C-terminal region of the protein.

A potential alternate explanation for the behavior of the histidine pairs is that the pore-interaction behavior might originate not from the histidines, but from their neighboring residues. However, this does not appear to be the case as the residues adjacent to the H3-H4, H10-H11 and H17-R18 pairs are all hydrophobic.

2.4 Discussion

In this work,we employ an approach to molecular dynamics simulation that allows the bilayer to self-assemble [7]. This has the advantage of preventing bias, but also prevents us from observing the early kinetic step in which the peptide binds the bilayer and begins to translocate. To be clear, in our system set-up, pores are formed even in the absence of peptide, and therefore our studies indicate details of peptide/bilayer interactions, rather than the kinetic process of peptide-induced pore formation. Nonetheless, in combination with multiple independent simulations of each system composition, this approach has the advantage of illustrating the variety of peptide/lipid structures with favorably low energy. The final peptide structures observed exhibit a wide variety of helicity (**Figure 2.5**), but a high degree of conservation of peptide-lipid interactions at the individual amino acid level (**Figure 2.6**). The simulations thus provide atomic-level insight into the structural plasticity that has long been observed experimentally for many helical AMPs and has been argued to

be a key part of their potency and selectivity [56, 58]. The observed structural heterogeneity relates to the concept of "imperfect amphipathicity" which suggests that AMP structures that present a few polar/charged residues on a non-polar face promote the formation/stabilization of pores [59,60]. In keeping with these ideas around AMP structural heterogeneity and imperfect amphipathicity, we observed that the most helical and perfectly amphipathic peptide structures (GAD-1-C and GAD-2-A) position parallel to the bilayer surface and tend to interact with the planar part of the bilayers formed. On the other hand, the majority of the simulations showed peptides with smaller helical contents interacting with the curved region of the bilayer in the pore. A variety of models have been employed in trying to understand AMP-induced pore formation, including the toroidal pore, carpet and barrel-stave models [3], and the two-state model [61], but the results with the Gad peptides are probably better understood in the context of models that capture the polymorphic/disordered characteristics of many AMPs, such as Bechinger et al. [56] and Sengupta et al. [57].

Another advantage of the self-assembly method is the appearance of a pore when the bilayer first forms. This allows us to determine whether and how the peptide interacts with a pore. The long lifetime of the pore, present even in the lipid-only simulations, is thus beneficial. While this persistence of the pore in our system is not unexpected, as explained by Fuertes et al. [62], it is significantly longer than the mere tens of nanoseconds reported in previous self-assembly studies [7, 63]. Repeating our self-assembly protocol for the lipid-only system with the SPC model of water, which was used in previous studies, instead of TIP4P, which we used in this work,we observed significantly faster bilayer formation and a pore lifetime of 50 ns. This consistency with previous work indicates that the choice of water model can have a

significant effect on membrane dynamics. This is not surprising given that seemingly small differences between water models can yield significantly different thermodynamic properties of simulated water [64]. We did examine other possible sources for this discrepancy including temperature, pressure control and other protocol details, and found no significant changes. Extending one of our lipid-only simulations (with TIP4P) reveals that the pore persists to at least 1 μs, and that the membrane is rather fluid, given a root-mean-square displacement of P atoms of approximately 2 nm in 100 ns. This confirms the view that the membrane-with-pore is a well-defined metastable state quite stable to thermal fluctuations, rather than a kinetically hindered state slowly and continuously moving towards equilibrium.

More specifically to histidine-containing AMPs, we observed a marked preference for the N-terminal half of GAD-1 and GAD-2 to interact more closely with the pore than the C-terminal half of the peptides (**Figure 2.7**). This preference corresponds well to the location of the sequential pairs of histidines, which are themselves much more closely positioned to the pores compared to other types of charged pairs such as histidine-arginine. There are substantial differences in structure between histidine and arginine (or lysine) side chains that might underlie the apparent differences in lipid interactions (**Figure 2.8**). Histidine represents a relatively compact, constrained side chain that, due to the two nitrogens present in the ring, has a hydrophilic nature in both its charged and uncharged states (**Figure 2.8-A**). In contrast, arginine consists of a long, conformationally flexible hydrophobic chain with a positively charged moiety at the terminus – and is thus most precisely viewed as amphipathic in character (**Figure 2.8-B**). Arginine has a well-known propensity to "snorkel" [65,66] i.e. embed its hydrophobic region in acyl chains of the bilayer and extend the charged terminal

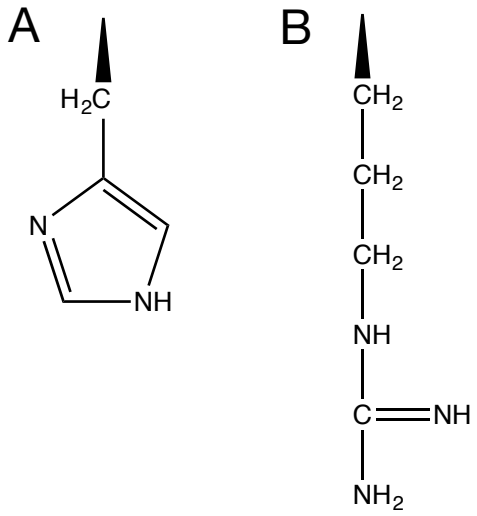

Figure 2.8: 2D chemical structure of A) Histidine and B) Arginine amino acids side chains. Histidine has a shorter and more rigid structure compared to arginine, for which the polar groups are located at the end of a hydrophobic CH_2 chain.

group out to the polar head group region. In comparison, a histidine-histidine pair presents a relatively conformationally constrained hydrophilic moiety that may not be able to interact easily with the densely packed lipid headgroups in the planar part of the bilayer. By contrast, the lipid headgroups in the pore are less tightly packed together and there is also more water available for favorable hydrophilic interactions and may thus provide a more energetically favorable site for the histidine pair to interact.

This potential explanation is consistent with the observation that even the uncharged histidine pairs associate with the pores and is interesting to consider in the light of studies which suggest that some histidine-containing AMPs have increased selectivity, even at neutral pH [22]. For example, Ruangsri et al. [5] have probed the antimicrobial activity of peptides with almost identical sequences to GAD-1 and

GAD-2 and suggested that while the GAD-1-like peptide has broad spectrum antimicrobial activity, the GAD-2-like peptide seems to be much more specific and was only found to be active against the fish parasite, Tetrahymena pyriformis. It is possible that the inclusion of histidine-pairs, perhaps in combination with the reduction in overall positive charge at neutral pH as in GAD-2 compared to GAD-1, presents one mechanism by which evolution may "tune" the structure-activity relationships of AMPs for specificity against particular types of pathogens. Note that, our observed absence of alteration in histidine-lipid interactions when histidine's charge is modified does not preclude a role for histidine charge in selectivity. It may well be the case that the more charged versions of the peptides bind more strongly to certain membranes, in particular to anionic ones, thus effecting membrane selectivity without necessarily affecting the mode of lipid interaction once it is bound.

In a final thought on the differences between GAD-1 and GAD-2, **Figure 2.9** schematically represents the two paralogs and the way they tend to interact with a pore in the bilayer. The H-H pairs prefer the more highly curved, central region of the pore where the increased spacing between head groups allows the compact, conformationally constrained pair of sidechains to interact favorably at the interface. On the other hand, the H-R pair is more able to interact with the more ordered, planar region. The basic premise of our simulations is that the structures we observe are representative of low free energy configurations and thus the differences in preferred positions suggest potentially interesting differences in the kinetics of how the pores are formed i.e., the GAD-1 preferred mode of interaction is suggestive of a more easily accessible initial defect-promoting step, whereas the GAD-2 preferred configuration is achieved more easily in the presence of a complete pore. The heterogeneity in the

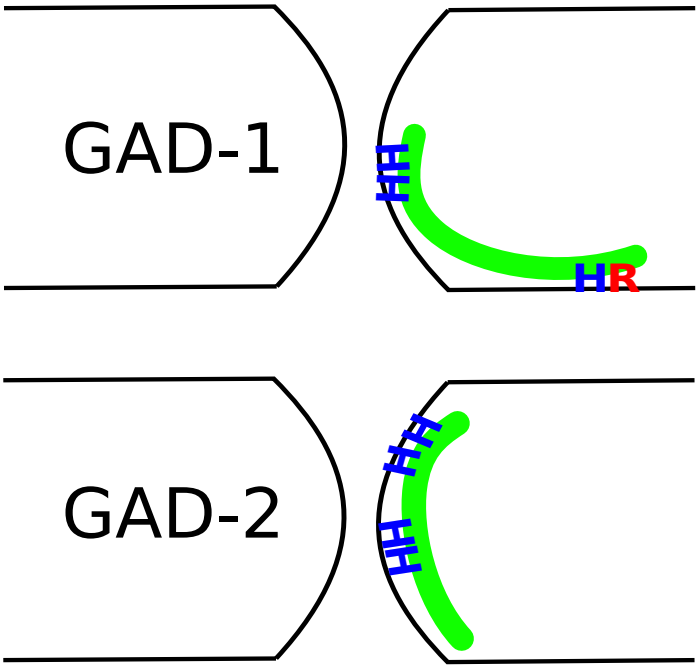

Figure 2.9: Schematic figure showing the two most favorable positions of GAD-1 and GAD-2 peptides.

final structures observed in our simulations suggests a kind of "stroboscopic" view of the potential kinetics of peptide promoted pore formation, with an initial interaction of the peptide with a planar bilayer looking like **Figure 2.4-C**, followed by the promotion of a defect in one leaflet as in **Figure 2.4-E** and then **Figure 2.4-A**, and perhaps translocation across the pore as in **Figure 2.4-B**. The model suggested by these simulations will be useful in the interpretation of experimental solution and solid-state NMR studies of GAD-1 and GAD-2 currently being performed in our lab. Another interesting issue to be addressed is the role of peptide oligomerization in pore interactions, as many AMPs are thought to form dimers, or higher order oligomers in order to exhibit their maximum toxicity [67–69]. Further studies employing more than one GAD peptide in the simulation box would help address this question.

Bibliography

[1] Wang, G., Li, X. and Wang, Z. "APD2: the updated antimicrobial peptide database and its application in peptide design" (2009) Nucleic Acids Res 37, D933-D937.

[2] Zasloff, M. "Antimicrobial peptides of multicellular organisms" (2002) Nature 415, 389-395.

[3] Brogden, K. A. "Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria?" (2005) Nature Reviews Microbiology 3, 238-250.

[4] Nguyen, L. T., Haney, E. F. and Vogel, H. J. "The expanding scope of antimicrobial peptide structures and their modes of action" (2011) Trends Biotechnol 29, 464-472.

[5] Bechinger, B. and Salnikov, E. S. "The expanding scope of antimicrobial peptide structures and their modes of action" (2012) Chem Phys Lipids 165, 282-301.

[6] Seo, M. D., Won, H. S., Kim, J. H., Mishig-Ochir, T., Lee, B. J. "Antimicrobial Peptides for Therapeutic Applications: A Review" (2012) Molecules 17, 12276-12286.

[7] Yeung, A. T., Gellatly, S. L., Hancock, R. E. "Multifunctional cationic host defence peptides and their clinical applications" (2011) Cell Mol Life Sci 68, 2161-2176.

[8] Rossi, L. M., Rangasamy, P., Zhang, J., Qiu, X. Q., Wu, G. Y. "Research advances in the development of peptide antibiotics" (2008) J Pharm Sci 97, 1060-1070.

[9] Hoskin, D. W., Ramamoorthy, A. "Studies on anticancer activities of antimicrobial peptides" (2008) Biochim Biophys Acta 1778, 357-375.

[10] Al-Benna,S., Shai, Y., Jacobsen, F., Steinstraesser, L. "Oncolytic Activities of Host Defense Peptides" (2011) Int J Mol Sci 12, 8027-8051.

[11] Yates, C., Sharp, S., Jones, J., Topps, D., Coleman, M., Aneja, R., Jaynes, J., Turner, T. "Oncolytic Activities of Host Defense Peptides" (2011) Biochem Pharmacol 81, 104-110.

[12] Grieco, P., Luca, V., Auriemma, L., Carotenuto, A., Saviello, M. R., Campiglia, P., Barra, D., Novellino, E., Mangoni, M. L. "Alanine scanning analysis and structure-function relationships of the frog-skin antimicrobial peptide temporin-1Ta" (2011) J Pept Sci 17, 358-365.

[13] Papo, N., Shai, Y. "New Lytic Peptides Based on the $_{D,L}$-Amphipathic Helix Motif Preferentially Kill Tumor Cells Compared to Normal Cells" (2003) Biochemistry 42, 9346-9354.

[14] Steinstraesser, L., Hauk, J., Schubert, C., Al-Benna, S., Stricker, I., Hatt, H., Shai, Y., Steinau, H. U., Jacobsen, F. "Suppression of Soft Tissue Sarcoma Growth by a Host Defense-Like Lytic Peptide" (2011) PLoS One 6, e18321.

[15] Rathinakumar, R., Wimley, W. C. "High-throughput discovery of broad-spectrum peptide antibiotics" (2010) FASEB J 24, 3232-3238.

[16] Kacprzyk, L., Rydengard, V., Morgelin, M., Davoudi, M., Pasupuleti, M., Malmsten, M. and Schmidtchen, A. "Antimicrobial activity of histidine-rich peptides is dependent on acidic conditions" (2007) Biochim Biophys Acta 1768, 2667-2680.

[17] Makovitzki, A., Fink, A. and Shai, Y. "Suppression of Human Solid Tumor Growth in Mice by Intratumor and Systemic Inoculation of Histidine-Rich and pH-Dependent Host Defense-like Lytic Peptides" (2009) Cancer Res 69, 3458-3463.

[18] Lee, I. H., Cho, Y. and Lehrer, R. I. "Effects of pH and salinity on the antimicrobial properties of clavanins" (1997) Infection and immunity 65, 2898-2903.

[19] Tu, Z., Young, A., Murphy, C., Liang, J. F. "The pH sensitivity of histidine-containing lytic peptides" (2009) J Pept Sci 15, 790-795.

[20] Chikakane, K., Takahashi, H. "Measurement of skin pH and its significance in cutaneous diseases" (1995) Clinics in dermatology 13, 299-306.

[21] Tannock, I. F. and Rotin, D. "Acid pH in tumors and its potential for therpeutic exploitation" (1989) Cancer research 49, 4373-4384.

[22] Kharidia, R., Tu, Z. , Chen, L., Liang, J. F. "Activity and selectivity of histidine-containing lytic peptides to antibiotic-resistant bacteria" (2012) Arch Microbiol 194, 769-778.

[23] Georgescu, J., Munhoz, V. H., Bechinger, B. "NMR Structures of the Histidine-Rich Peptide LAH4 in Micellar Environments: Membrane Insertion, pH-Dependent Mode of Antimicrobial Action, and DNA Transfection" (2010) Biophys J 99, 2507-2515.

[24] Browne, M. J., Feng, C. Y., Booth, V. and Rise, M. L. "Characterization and expression studies of Gaduscidin-1 and Gaduscidin-2; paralogous antimicrobial peptide-like transcripts from Atlantic cod (Gadus morhua)" (2011) Dev Comp Immunol 35, 399-408.

[25] Ruangsri, J., Fernandes, J. M. O., Rombout, J. H. W. M., Brinchmann, M. F. and Kiron, V. "Ubiquitous presence of piscidin-1 in Atlantic cod as evidenced by immunolocalisation" (2012) BMC veterinary research 8, 46.

[26] Ruangsri, J., Salger, S. A., Caipang, C., Kiron, V. and Fernandes, J. M. O. "Differential expression and biological activity of two piscidin paralogues and a novel splice variant in Atlantic cod (Gadus morhua L.)" (2012) Fish & Shellfish Immunology 32, 396-406.

[27] Fernandes, J. M. O., Ruangsri, J. and Kiron, V. "Atlantic Cod Piscidin and Its Diversification through Positive Selection" (2010) Plos one 5, e9501.

[28] Campagna, S., Saint, N., Molle, G., Aumelas, A. "Structure and Mechanism of Action of the Antimicrobial Peptide Piscidin" (2007) Biochemistry 46, 1771-1778.

[29] Chekmenev, E. Y., Jones, S. M., Nikolayeva, Y. N., Vollmar, B. S., Wagner, T. J., Gor'kov, P. L., Brey, W. W., Manion, M. N., Daugherty, K. C., Cotten, M. "High-Field NMR Studies of Molecular Recognition and Structure-Function Relationships in Antimicrobial Piscidins at the Water-Lipid Bilayer Interface" (2006) J Am Chem Soc 128, 5308-5309.

[30] Chekmenev, E. Y., Vollmar, B. S., Forseth, K. T., Manion, M. N., Jones, S. M., Wagner, T. J., Endicott, R. M., Kyriss, B. P., Homem, L. M., Pate, M., He, J., Raines, J., Gor'kov, P. L., Brey, W. W., Mitchell, D., J., Auman, A. J., Ellard-Ivey, M. J., Blazyk, J., Cotten, M. "Investigating molecular recognition and biological function at interfaces using piscidins, antimicrobial peptides from fish" (2006) BiochimBiophys Acta 1758, 1359-1372.

[31] De Angelis, A. A., Grant, C. V., Baxter, M. K., McGavin, J. A., Opella, S. J., Cotten, M. L. "Amphipathic Antimicrobial Piscidin in Magnetically Aligned Lipid B" (2011) Biophys J 101, 1086-1094.

[32] Fu, R., Gordon, E., D., Hibbard, D. J., Cotten, M. "High Resolution Heteronuclear Correlation NMR Spectroscopy of an Antimicrobial Peptide in Aligned Lipid Bilayers: Peptide-Water Interactions at the Water-Bilayer Interface" (2009) J Am Chem Soc 131, 10830-10831.

[33] Lee, S. A., Kim, Y. K., Lim, S. S., Zhu, W. L., Ko, H., Shin, S. Y., Hahm, K. S., Kim, Y. "Solution Structure and Cell Selectivity of Piscidin 1 and Its Analogues" (2007) Biochemistry 46, 3653-3663.

[34] Lin, H. J., Huang, T. C., Muthusamy, S., Lee, J. F. Duann, Y. F., Lin, C. H. "Piscidin-1, an Antimicrobial Peptide from Fish (Hybrid Striped Bass Morone saxatilis x M. chrysops), Induces Apoptotic and Necrotic Activity in HT1080 Cells" (2012) Zoolog Sci 29, 327-332.

[35] Mehrnejad, F., Zarei, M. "Molecular Dynamics Simulation Study of the Interaction of Piscidin 1 with DPPC Bilayers: Structure-Activity Relationship" (2010) J Biomol Struct Dyn 27(4), 551-560.

[36] Niu, S. F., Jin, Y., Xu, X., Qiao, Y., Wu, Y., Mao, Y., Su, Y. Q., Wang, J. "Characterization of a novel piscidin-like antimicrobial peptide from Pseudosciaena crocea and its immune response to Cryptocaryon irritans" (2013) Fish Shellfish Immunol 35, 513-524.

[37] Yuan, T., Zhang, X., Hu, Z., Wang, F., Lei, M. "Molecular dynamics studies of the antimicrobial peptides piscidin 1 and its mutants with a DOPC lipid bilayer" (2012) Biopolymers 97, (2012) 998-1009.

[38] Esteban-Martín, S. and Salgado, J. "Self-Assembling of Peptide/Membrane Complexes by Atomistic Molecular Dynamics Simulations" (2007) Biophysical journal 92, 903-912.

[39] Leontiadou, H., Mark, A. E., Marrink, S. J. "Antimicrobial Peptides in Action" (2006) J Am Chem Soc 128, 12156-12161.

[40] Ulmschneider, J. P., Doux, J. P. F., Killian, A. J., Smith, J. C., Ulmschneider, M. B. "Peptide Partitioning and Folding into Lipid Bilayers" (2009) J. Chem. Theory Comput. 5, 2202-2205.

[41] Arnold, K., Bordoli, L., Kopp, J., Schwede, T. "The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling" (2006) Bioinformatics 22, 195-201.

[42] Kiefer, F., Arnold, K., Kunzli, M., Bordoli, L., Schwede, T. "The SWISS-MODEL Repository and associated resources" (2009) Nucleic Acids Res 37, D387-D392.

[43] Peitsch, M. C. "From amino acid sequence to protein structure: A free one-hour service" (1995) Bio-Technology 13, 658-660.

[44] Swiss-PdbViewer webpage:

http://www.expasy.org/spdbv/ (accessed: 09/27/2015).

[45] Zhao, W., Rog, T., Gurtovenko, A. A., Vattulainen, I. and Karttunen, M. "Atomic-Scale Structure and Electrostatics of Anionic Palmitoyloleoylphosphatidylglycerol Lipid Bilayers with Na^+ Counterions " (2007) Biophys J 92, 1114-1124.

[46] Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M. and van der Spoel, D. "GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit" (2013) Bioinformatics 29, 845-854.

[47] Jorgensen, W. L., Maxwell, D. S., Tirado-Rives, J. "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids" (1996) Journal of the American Chemical Society 118, 11225-11236.

[48] Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., Jorgensen, W. L. "Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides" (2001) J. Phys. Chem. B 105, 6474-6487.

[49] Tieleman, D. P., Maccallum,J. L., Ash, W. L., Kandt, C., Xu, Z., Monticelli, L. "Membrane protein simulations with a united-atom lipid and all-atom protein model: lipid?protein interactions, side chain transfer free energies and model proteins" (2006) J Phys Condens Matter 18, S1221-S1234.

[50] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L. "Comparison of simple potential functions for simulating liquid water" (1983) J. Chem. Phys. 79, 926.

[51] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., Hermans, J. "Interaction Models for Water in Relation to Protein Hydration" (1981) In: Pullman B (ed) Intermolecular forces. Reidel, Dordrecht 14, pp 331-342.

[52] Kabsch, W. and Sander, C. "Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features" (1983) Biopolymers 22, 2577-2637.

[53] Joosten, R. P., Te Beek, T. A. H., Krieger, E., Hekkelman, M. L., Hooft, R. W. W., Schneider, R., Sander, C. and Vriend, G. "A series of PDB related databases for everyday needs" (2011) Nucleic acids research 39, D411-D419.

[54] Humphrey, W., Dalke, A. and Schulten, K. " : Visual molecular dynamics" (1996) Journal of molecular graphics 14, 33-38.

[55] VMD molecular visualization program introduction webpage: http://www.ks.uiuc.edu/Research/vmd/ (accessed: 02/27/2014).

[56] Bechinger, B. and Aisenbrey, C. "The Polymorphic Nature of Membrane-Active Peptides from Biophysical and Structural Investigations" (2012) Current Protein and Peptide Science, 13, 602-610.

[57] Sengupta, D. , Leontiadou, H., Mark, A. E., Marrink, S.J. "Toroidal pores formed by antimicrobial peptides show significant disorder" (2008) Biochim Biophys Acta 1778, 2308-2317.

[58] Vermeer, L. S., Lan, Y., Abbate, V., Ruh, E., Bui, T. T., Wilkinson, L. J., Kanno, T., Jumagulova, E., Kozlowska, J., Patel, J., McIntyre, C. A., Yam, W. C., Siu, G., Atkinson, R. A., Lam, J. K., Bansal, S. S., Drake, A. F., Mitchell, G. H. and Mason, A. J. "Conformational Flexibility Determines Selectivity and Antibacterial, Antiplasmodial, and Anticancer Potency of Cationic α-Helical Peptides" (2012) J Biol Chem 287, 34120-34133..

[59] Mihajlovic, M. and Lazaridis, T. "Charge distribution and imperfect amphipathicity affect pore formation by antimicrobial peptides" (2010) Biochim Biophys Acta 1818, 1274-1283.

[60] Mihajlovic, M. and Lazaridis, T. "Antimicrobial peptides bind more strongly to membrane pores" (2010) Biochim Biophys Acta 1798, 1494-1502.

[61] Chen, F. Y., Lee, M. T., Huang, H. W. "Evidence for Membrane Thinning Effect as the Mechanism for Peptide-Induced Pore Formation" (2003) Biophysical journal 84, 3751-3758.

[62] Fuertes, G., Gimenez, D., Esteban-Martin, S., Sanchez-Munoz, O. L., Salgado, J. "A lipocentric view of peptide-induced pores" (2011) Eur Biophys J 40, 399-415.

[63] Marrink, S. J., Lindahl, E., Edholm, O., Mark, A. E. "Simulation of the Spontaneous Aggregation of Phospholipids into Bilayers" (2001) J. Am. Chem. Soc. 123, 8638-8639.

[64] Vega, C., Abascal, J. L. F., Sanz, E., MacDowell, L.G., McBride, C. "Can simple models describe the phase diagram of water?" (2005) J. Phys.: Condens. Matter 17, S3283-S3288.

[65] Mishra, V. K., Palgunachari, M. N., Segrest, J. P. and Anantharamaiah, G. M. "Interactions of synthetic peptide analogs of the class A amphipathic helix with lipids. Evidence for the snorkel hypothesis" (1994) Journal of Biological Chemistry 269, 7185-7191.

[66] Hristova, K., Wimley, W. C. "A Look at Arginine in Membranes" (2011) The Journal of membrane biology 239, 49-56.

[67] Yeaman, M. R., Yount, N. Y. "Mechanisms of Antimicrobial Peptide Action and Resistance" (2003) Pharmacol Rev 55, 27-55.

[68] Oren, Z., Lerman, J. C., Gudmundsson, G. H., Agerberth, B., Shai, Y. "Structure and organization of the human antimicrobial peptide LL-37 in phospholipid membranes: relevance to the molecular basis for its non-cell-selective activity" (1999) Biochem J 341, 501-513.

[69] Silvestro, L., Axelsen, P. H. "Membrane-Induced Folding of Cecropin A" (2000) Biophysical Journal 79, 1465-1477.

Chapter 3

Computational studies of pulmonary surfactant protein SP-B interacting with lipid bilayers[1]

3.1 Introduction

By weight, human lung surfactant (LS) is a mixture of 80-85% phospholipids, 5-10% neutral lipids and 10% proteins [1–3]. About 80% of the phospholipids are phosphocholine (PC) lipids, half of which are dipalmitoylphosphatidylcholine (DPPC). Along with PC lipids, a significant amount of phosphatidylglycerol (PG) lipid is in LS. The protein part of lung surfactant is composed of pulmonary surfactant proteins SP-A, SP-B, SP-C and SP-D [4]. Hydrophilic surfactant proteins (SP-A and

[1] A manuscript based on the all-atom simulations in this chapter has been accepted for publication in Biophysica et Biochimica Acta (BBA)-Biomembranes (2016).

SP-D) have antimicrobial activity and are responsible for immune functions within alveoli [5,6], while the hydrophobic surfactant proteins (SP-B and SP-C) increase the surface activity of LS lipids, easing the breathing process [7,8]. Additionally, the work Robertson and coworkers demonstrated antimicrobial activity of SP-B and SP-C on Group B streptococci pneumonia in rabbit lung [9], and on Group B streptococci, Staphylococcus aureus and Escherichia coli bacteria in vitro [10].

Previous studies provide a list of *in vitro* functions for SP-B: membrane binding, membrane lysis, membrane fusion, promotion of lipid adsorption to air-liquid intersurfaces, stabilization of monomolecular surface films, and re-spreading of films from a collapsed phase [8]. Moreover, it has been shown that the hydrophobic surfactant proteins (SP-B and SP-C) can form pores in artificial lipid bilayers to transport cations [11]. Despite its importance for life, the structure of SP-B at atomic resolution is unknown and consequently a detailed understanding of its mechanism for how it performs these activities is lacking. In this part of my thesis the goal is to find feasible low energy structures and topologies for SP-B that will help reveal how SP-B binds to a membrane and interacts with lipid bilayers. Structures obtained from simulation will be useful to interpreting experimental studies of SP-B structure.

SP-B is a 79 residue protein from the saposin super family. Its sequence is FPIP LPYCWL CRALIKRIQA MIPKGALAVA VAQVCRVVPL VAGGICQCLA ERYSV ILLDT LLGRMLPQLV CRLVLRCSMD [8, 12] and is shown in **Figure 3.1**, along with that of Mini-B [17], a shorter construct of SP-B with known experimental structure. Functionally, the first 7 N-terminal residues are believed to be important to promote insertion of SP-B from water into the air/water interface [15, 16, 22]. SP-B also contains 3 disulphide bonds between cysteines C8-C77, C11-C71 and C35-C46 [17].

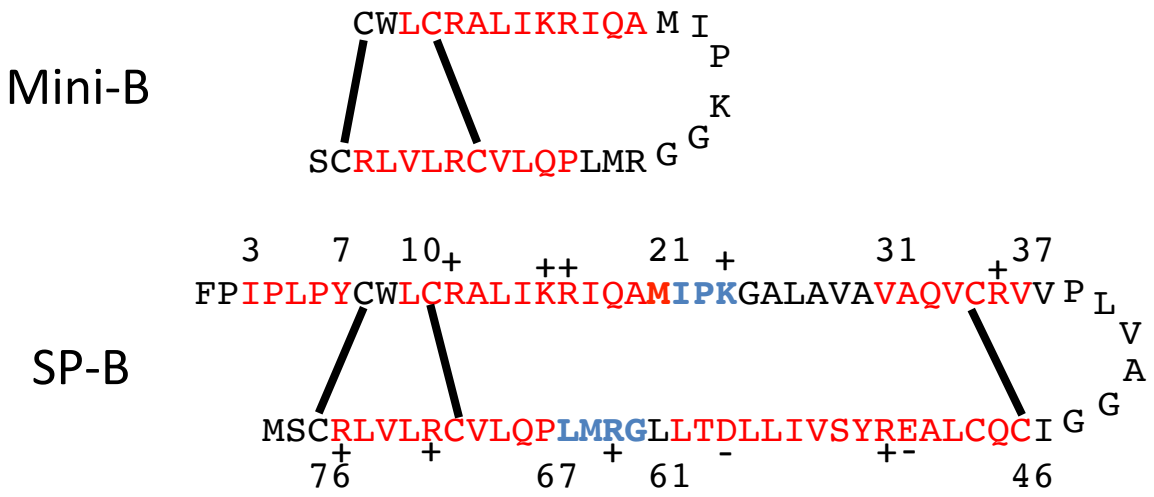

Figure 3.1: Amino acid sequence of SP-B and Mini-B. Red residues indicate helical regions, known for Mini-B and predicted for SP-B (all in α-helical structure, except for C46-G47-C48, which are in 3_{10} helix structure). We can predict that SP-B has the same N and C-teminal α-helical regions as the corresponding regions in Mini-B. Blue-bolded residues indicate regions assumed to allow for SP-B bending. M-21 is helical in the initial open structure and non-helical in the initial bent structure. Heavy lines indicate disulphide bonds. Details of how the secondary structure was predicted can be found in the Methods section.

Owing to SP-B's exceptional degree of hydrophobicity, experimental studies of SP-B's structure are extremely difficult and have yet to yield a high resolution structure. SP-B has a GRAVY index of 1.027 (**Table 3.1**) [26]. The GRAVY index indicates the hydrophobicity of a protein: the more hydrophobic the protein, the higher the GRAVY index (**Table 3.1**).

Experimental structures of proteins from the saposin super family, including Saposin A [5], Saposin B [6], Saposin C [21–23], NK-lysin [19], granulysin [7] and amoeba-

Protein	Gravy Index
GNLY_HUM	-0.648
NKL_PIG	-0.117
SAP_A_HUM	0.007
SAP_B_HUM	0.001
SAP_C_HUM	0.074
SAP_D_HUM	0.082
SP-B_HUM	1.027
AMP_A_ENT	0.406

Table 3.1: Gravy index of SP-B and the other proteins in the saposin family. We can see that the GRAVY index of SP-B is very high compared with the rest of the proteins, indicating the extreme hydrophobicity of SP-B. The GRAVY values calculate using an online GRAVY index calculator (http://www.bioinformatics.org/sms2/protein_gravy.html) [67].

pore A [8, 26], are similar to each other in displaying 4 or 5 helical regions in their secondary structure, as well as three internal disulphide bonds. Although their secondary structure is similar, in their tertiary structure they have a variety of structures (**Figure 3.2**). They can be in an open structure (**Figure 3.2-A**), bent (V-shape) structure (**Figure 3.2-B**) or closed structure (**Figure 3.2-C**).

To carry out simulations on systems containing a protein and a large number of lipid and water molecules, we should keep in mind that performing all-atom (AA) force field simulations will give us more realistic and detailed structures at the cost of

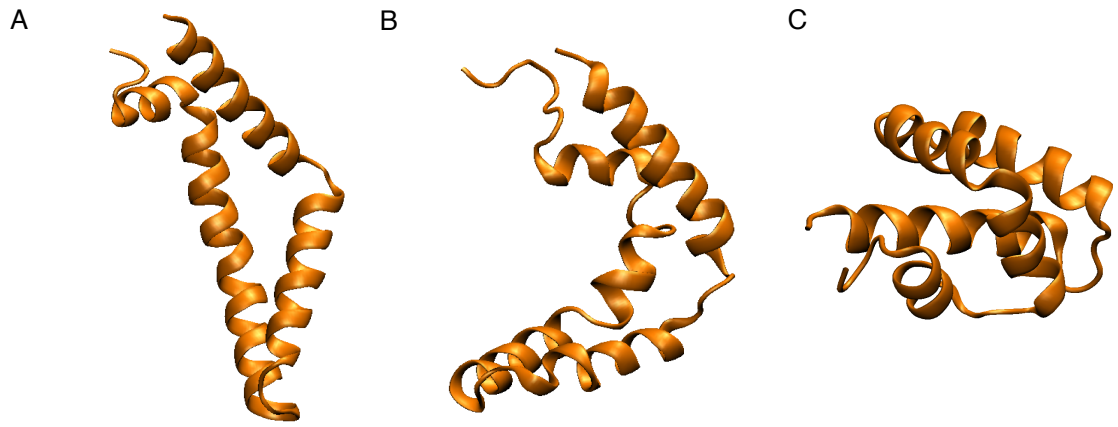

Figure 3.2: Different tertiary structures of the saposin family, A) Saposin A (PDB ID: 4DDJ) in an open structure, B) Saposin C (PDB ID: 1N69) in a bent structure, C) NK-lysin (PDB ID: 1NKL) in a closed structure. The structure of Saposin A (in LDAO detergent) and Saposin C (in a buffer) are determined using X-ray crystallography, while the structure of NK-lysin (in water) is determined using solution NMR.

consuming significant computer time. By using coarse-grained (CG) force fields, the simulations will run faster, but we will lose the atomistic resolution of the system. Recent advances in computation hardware and software have made it possible to simulate large systems. A recently developed coarse-grained force field is the MARTINI model [24–26], which is now becoming widely used for large biomolecules [23,27,36,37]. To increase the accuracy of the simulations, researchers use hybrid model force fields in which they use different force fields for different components of the simulation, depending on the relative importance of the system components. Such a scheme may employ the combination of an AA force field or united-atom (UA) force field for proteins in the system and a CG force field for the rest of the molecules in the system,

e.g., water and lipids [17, 38–40]. Using such an approach helps the simulation run faster while retaining more detail on system components of interest.

Several researchers have employed modelling approaches to carry out simulations to predict the structure and behaviour of full-length SP-B in lipid bilayers. Zaltash et al. [42] have studied SP-B's monomeric and dimeric structures using NK-lysin as a template for the secondary structure of the SP-B using Monte Carlo simulations. They have also used polymyxin-B as a template to study the functionality of protein in lipid bilayer (polymyxin-B mimics SP-B's functions but it is structurally different). Based on their model, they have predicted that the distribution of polar and non-polar residues in the SP-B dimer is compatible both with SP-B positioning at the water-lipid interface of a bilayer and with the ability of SP-B to cross-link membrane multilayers. Baoukina and coworkers have performed coarse grained simulations on monomers of SP-B and SP-C interacting with vesicles. They have studied the fusion of two lipid vesicles in the presence of SP-B monomers [23] and the fusion of lipid vesicles and bicelles into lipid monolayers in the presence of SP-B monomers and dimers [27]. They have modelled SP-B employing homology modelling based on NK-lysin, Saposin C and Mini-B structures. In their studies they have seen SP-B present as both open and closed structures, which will be discussed later in this chapter.

SP-B purified from animal lungs exhibits a variety of oligomeric states and deletion of C48 does not impair SP-B's function [43]. It has been suggested that, in addition to the C48-C48 disulfide, SP-B dimers may be stabilized through salt bridges between E51 and R52 [42, 44]. Olmeda et al. [27] have recently proposed that SP-B acts as a multimer of dimers, forming hydrophobic ring/tube shapes. Thus, SP-B's functional oligomerization state is still something of an open question.

In this chapter we are trying to find low energy 3D structures of SP-B using computer simulations, employing both MD and REMD techniques. To reach this goal one should run simulations for a sufficiently long time for the protein to find the low energy structures. I run most of our all-atom simulations for ≥ 2 μs, which appears to be long enough to approach equilibrium, in that we see some convergence of structure from different starting configurations and large-scale conformational changes (**Table 3.4**). These simulations give us detailed information about the structure of SP-B and ideas about the functional mechanism of the protein. To further explore the structures, we employ a hybrid force field, PACE, to run two sets of MD simulations, as well as a series of REMD simulations. PACE uses a UA force field for the protein part and the MARTINI force field for the rest of the particles in the simulation box [17, 18]. To reduce the computer time required to equilibrate the protein, we try to use reasonable starting model structures based on homology modelling, experimentally determined structures for smaller proteins derived from SP-B and ideas from the literature on the secondary structure of SP-B.

3.2 Methods

3.2.1 Modelling the Initial Structure of SP-B

A useful method to predict the structure of protein is homology modelling, which predicts the 3D structure of a protein with an unknown structure based on sequence similarity to a protein of known structure. Because SP-B is from the saposin family of proteins, we can use the known structures of proteins from this protein family

to carry out homology modelling of SP-B. This method is challenging because the saposins with known structure are not very similar in sequence to SP-B.

Running the basic local alignment search tool (BLAST) [31] on the UniPort webpage (http://www.uniprot.org/blast/) indicates roughly near 25% identity (for Saposin C) in sequence of SP-B as the maximum compared to other proteins in the saposin super family. The homology model based on one of the Saposin C structures (PDB IDs: 2QYP [21]) produces 82% helicity, which is much higher than the experimental values for SP-B (35-50% [26]). Homology modelling based on the other Saposin C structures (1SN6 [22] and 1M12 [23]), along with the problem of high percent helicity, the model produced does not form the third disulfide bond (C35-C46).

Our approach is to pair information from homology modelling based on both NK-lysin and Saposin C with the experimentally-determined structure of Mini-B, a construct derived from SP-B; The structure of portions of SP-B that are in common with Mini-B are given on the structure of Mini-B, while the structure of the rest of SP-B is inferred from homology to NK-lysin and Saposin C with the idea of the SP-B having amphipathic helical regions. Residues 26-62 of SP-B (residues not included in Mini-B) do not have residue-specific experimental data, but, based on secondary structure prediction techniques applied to SP-B [2,22] and homology to other saposins, are believed to contain two or three amphipathic α-helices. In **Table 3.2** we see a comparison of helical regions in our model (**Figure 3.1**) to some other models introduced for SP-B [23, 27, 48, 49]. Given the large proportion of proline in the first seven residues (F-P-I-P-L-P-Y), this initial sequence is likely to take on an extended, poly-proline helix-like structure. We produce our SP-B sequence by keeping the helical regions of Mini-B (which we call H-II and H-V), and introducing helical structure to

Ile-3–Tyr-7 (H-I), Val-31–Val-37 (H-III) and Cys-46–Leu-61 (H-IV) (**Figure 3.1**) and (**Table 3.2**). Helices II-V are predicted to be α-helical based on homology (C46-G47-C48 are in the 3_{10} helical structure). We employed both STRIDE [42] as a VMD plugin and DSSP [51, 52] as a GROMACS tool, to calculate the secondary structure of the protein. These methods use two different ways of calculating the secondary structure. Although employing both methods to calculate the helical structure of the initial structures of SP-B gives similar results, there are differences between them, e. g. in the initial structure of SP-B C46-G47-C48 were defined to be in 3_{10} helical structure in STRIDE plug-in of VMD while the DSSP plug-in of GROMACS called this region α-helical. Thus, we call all the three secondary structures of α-helical, 3_{10}-helical and π-helical (which rarely occurs in simulations) with the general name of "helical" structure. Residues 1-7 are likely to be in an extended, possibly poly-proline helix like structure. As an artifact of energy minimization of the structure in SPV software, which we used to build the structures, residues 3-7 were given an alpha-helical starting structure.

We make our SP-B protein in two different structures, open and bent (V-shape) (**Figure 3.3**). The regions at the bend are chosen to be Met-21–Lys-23 and Gly-63–Leu-66 in order to preserve the helical regions (**Figure 3.1**). The whole procedure of constructing our SP-B model is done using Swiss PDB viewer (SPV) software [53].

For our simulations, we do not construct a model of the protein in a closed structure. Adding new structures increases the number of simulations, and hence the demand for computational resources, required to validate them. Rather, we wish to determine whether SP-B tends to have an open or closed structure, and do so by simulating the bent structure, which could either open up or close in, depending on

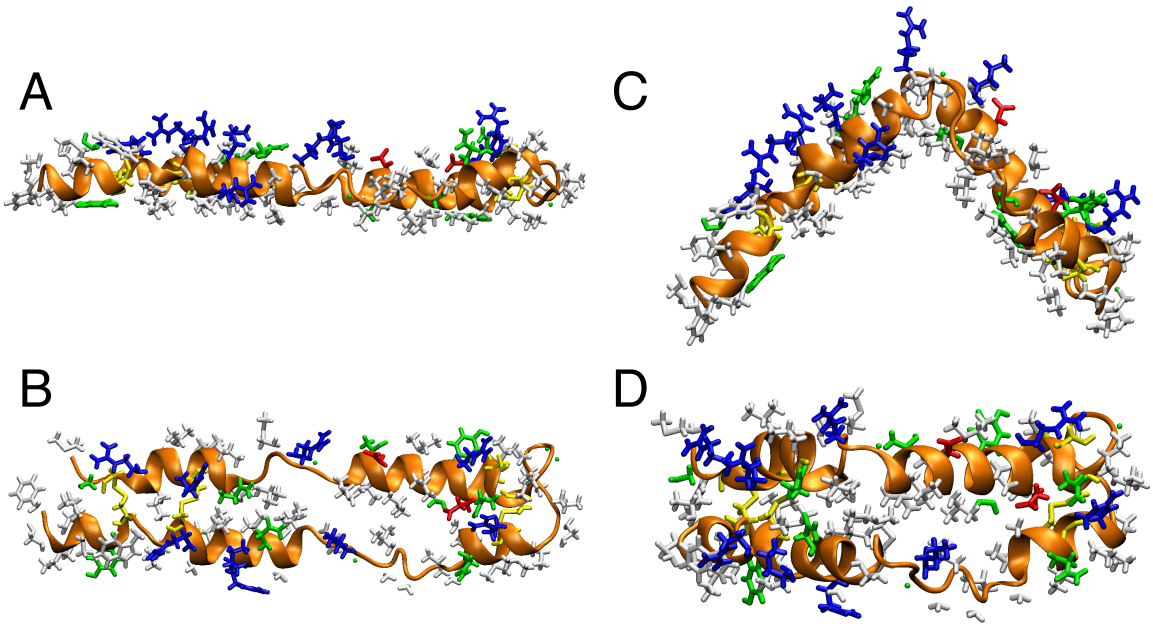

Figure 3.3: The two starting structure models of SP-B based on Mini-B structure (PDB ID: 2DWF) and homology to saposin super family structures. Here we can see the open structure of SP-B: A) side view and B) top view; and the bent (V-shape) structure of SP-B: C) side view and D) top view. The protein backbone is shown as an orange ribbon, the cysteine sidechains in yellow, hydrophobic sidechains in white, polar uncharged sidechains in green, positively charged sidechains in blue, and negatively charged sidechains in red.

the energetics of the protein.

3.2.1.1 Molecular Dynamics

In carrying out MD simulations we use both OPLS-AA and PACE force fields. We try to keep the initial system configurations of the simulations the same for both force fields (**Table 3.3**). In setting up MD simulations that use OPLS-AA, we use a pre-assembled POPC bilayer composed of 512 lipid molecules. The bilayer employed has a pore in order to allow lipid molecules to move from one leaflet to the other in order to keep the lipid/area ratio in the whole bilayer constant. To generate the

Helical region	open (bent) AA model	open (bent) UA model	model-A	model-B	model-C
helix-I	3-7 (3-7)	3-7 (3-7)	-	-	-
helix-II	10-21 (10-20)	10-21 (10-21)	7-21	10-20	8-22
helix-III	31-37 (31-37)	30-37 (30-37)	32-36	26-35	27-38
helix-IV	46-61 (46-62)	46-61 (46-62)	42-64	44-58	42-50
helix-V	67-76 (67-76)	67-76 (67-76)	68-76	67-78	67-74

Table 3.2: Proposed helical regions, given by residue number, of SP-B protein in different models. The open(bent) structure for the all-atom force field and open(bent) structure for the united-atom force field are the four models we use in our simulations. The previously used models are Model-A, introduced by Tieleman and coworkers [23, 27], Model-B, which was a model previously used in our group [48], and model-C, introduced by the Johansson group [49]. The helical regions for bent structures are enclosed in parentheses.

bilayer, we start with 3 pre-assembled perfect lipid bilayers, each composed of 128 lipid molecules and one lipid bilayer containing a pore, also composed of 128 POPC lipids (from one of our control simulations in the previous chapter). We put these four bilayer sections beside each other in a square to produce our lipid bilayer containing 512 POPC lipid molecules. To equilibrate the new lipid bilayer, we add water to the system and run it under NVT conditions for 2 ns at 310 K, and subsequently under NPT conditions for 50 ns at 310 K. The pore placed in the membrane remains in the bilayer during the simulation as we discussed in Chapter 2.

	Packages	
Box Contents	GROMACS 4.5.5	GROMACS 4.6.3
Force Field	OPLS-AA	PACE
Method	MD	MD & REMD
Number of POPC	496	~ 400
Number of Water	38000	~ 11000
Number of Cl ions	7	7
Number of SP-B	1	1

Table 3.3: Contents of simulation box for our simulations.

Then we add protein to the system. For each of the structures (open and bent) we put the protein in three different positions with respect to the bilayer: inside (in), half inside (half) and outside (out) the bilayer (**Figure 3.4**, **Figure 3.5** and **Figure 3.6**). To put the protein in a more favourable position, we put the hydrophobic face of the protein pointing towards the bilayer centre and hydrophilic face towards the water. SP-B is placed far from the (preformed) pore to prevent the pore and protein from interacting. To put the protein inside the bilayer, we use the g_membed tool in GROMACS, which embeds protein in the lipid bilayer by removing the least number of lipid molecules. Each of our six simulation boxes generated in this way has 496 POPC lipid molecules, 38000 TIP4P water molecules, 7 Cl^- counter ions and an SP-B protein (**Table 3.3**). The system size in x, y and z directions before starting the simulation are ~ 12.4 nm, ~ 12.3 nm and ~ 11.7 nm, respectively.

For each of the simulation boxes we use the same method to set up and run the

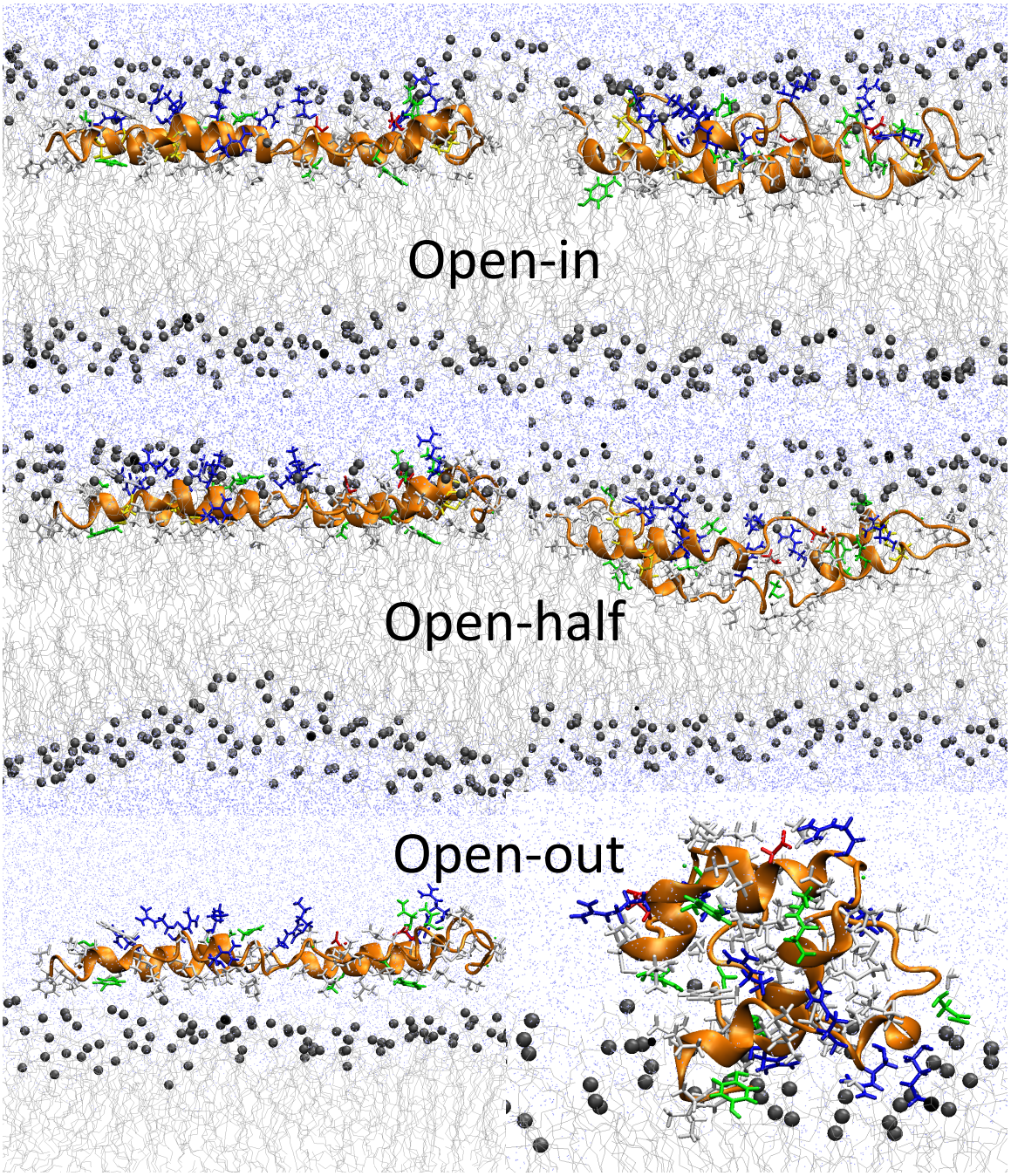

Figure 3.4: Initial and final positions of the open structure shown in relation to the lipid bilayer in all-atom simulations. Left: different initial positions. Right: The corresponding final structure at the end of the simulation.

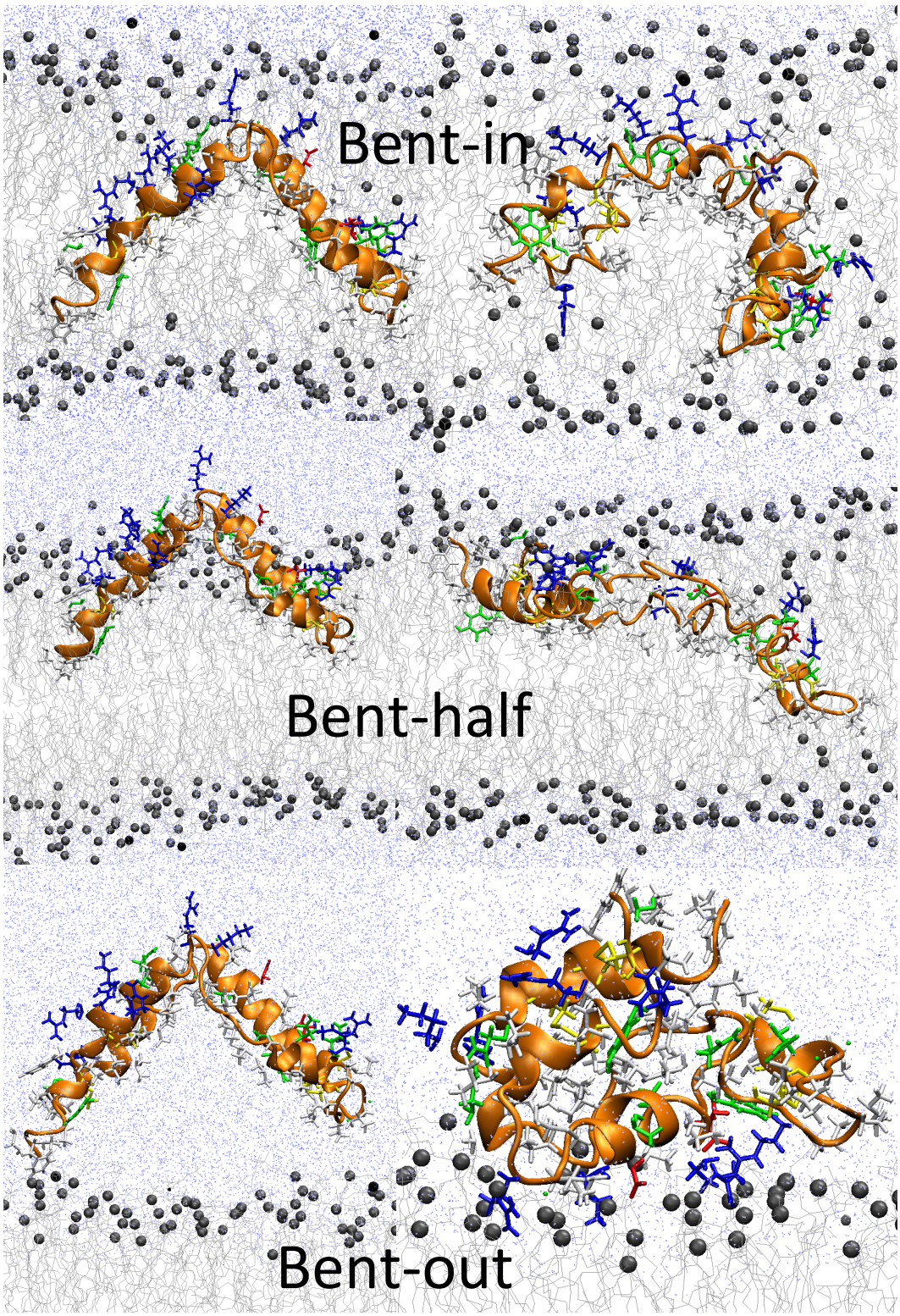

Figure 3.5: Initial and final positions of the bent structure viewed in relation to the lipid bilayer in all-atom simulations. Left: different initial positions. Right: The corresponding final structure at the end of the simulation.

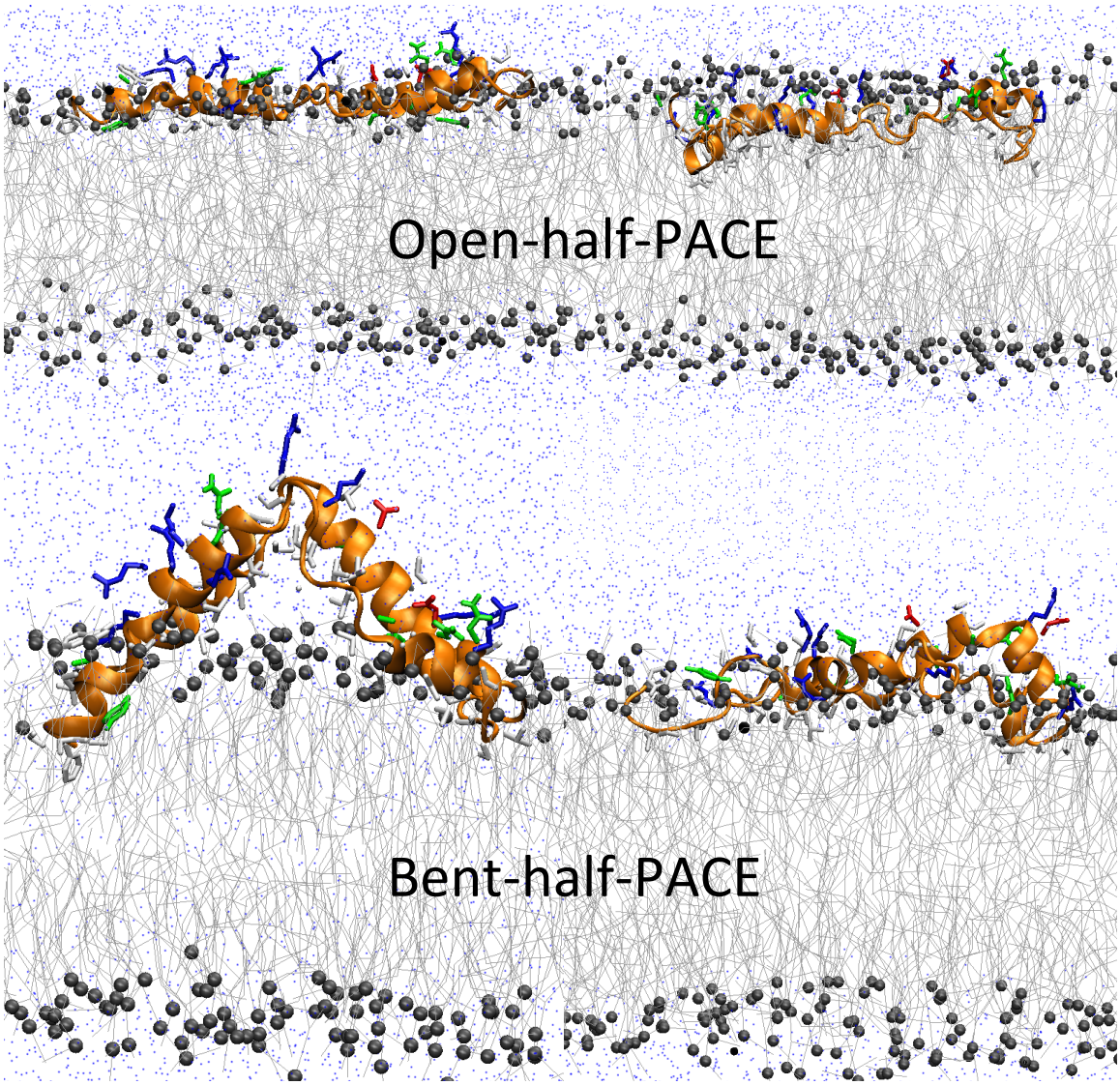

Figure 3.6: Initial and final positions of Mini-B-based open with respect to the lipid bilayer in PACE simulations. Left: Different initial positions of open with respect to the lipid bilayer. Right: The corresponding final structure of open at the end of simulation.

simulations. In the first step, under NVT conditions we run the system at 310 K for 2 ns with protein position restrained and with a time step of 2 fs. We keep the protein position restrained to prevent any structural changes before the production run. In the second step, we run the system under NPT condition at 310 K and 1 atm for 100 ps with protein position restrained and using Parrinello-Rahman isotropic pressure coupling with $\tau_p = 5$ ps and compressibility = 4.5×10^{-5} bar^{-1}. In the final step, the production run, we simulate the system under NPT conditions at 310 K and 1 atm, using Parrinello-Rahman semi-isotropic pressure coupling, with the same time constant and compressibility as before and removing the protein position restraints. We run these simulations with a time step of 2 fs for 2 to 2.5 μs in duration, except for the bent-out simulation, which runs for 700 ns as the protein fairly quickly forms the closed structure and remains closed. Details of the simulation parameters are in Appendix A.

In addition to the OPLS-AA all-atom force field, we use the PACE force field. We try to keep the initial setup of the PACE simulations as close as possible to the OPLS-AA initial setup. To set up the simulation box and to embed the protein in lipid bilayer, we use a web based protocol explained in Ref. (58), developed for this force field. We upload the PDB file of our SP-B to the webpage [59]. Then we select the type and size of the bilayer and the distance and orientation of the protein with respect to the prepared bilayer. In our case, we produce a POPC bilayer containing $\sim$ 400 POPC molecules. After that, we place the bent and open forms of the protein partially inside the bilayer with the hydrophobic side facing toward the bilayer centre. Then we add PACE water molecules (MARTINI water, where 4 water molecules are represented as a single coarse-grained bead) and neutralize the system by adding 7

Cl^- ions (MARTINI model). The system dimensions in the x, y and z directions before starting the simulation are $\sim$ 13 nm, $\sim$ 13 nm and $\sim$ 9 nm, respectively.

As a cautionary note, we mention that for MARTINI water molecules, the GROMACS utility that normally solvates the system with more realistic models of water, adds what is in effect too many water molecules, leading to unwanted crystallization owing to a high density. To prevent this, we simply add $\sim$20% fewer water molecules, so that they remain in the liquid state. We add $\sim$ 11000 water "beads" to fill the simulation box.

We run the system under NVT conditions at T = 310 K for 2 ns with a time step of 5 fs, restraining the position of the protein. In the second and final step, we start the production run. Under NPT condition using Parrinello-Rahman, semi-isotropic pressure coupling with τ_p = 5 ps and compressibility = 4.5×10^{-5} bar^{-1}, at 310 K, we run the system for 3 μs with a time step of 5 fs, without position restraints on the protein. Details of the simulation parameters are in Appendix A. Based on the initial protein position and conformation, we call the two PACE simulations "bent-half-PACE" and "open-half-PACE".

3.2.1.2 REMD

We carry out replica exchange molecular dynamics only with the PACE force field as all-atom simulations would be prohibitively costly to run. The temperatures we use range from 300 K up to 370 K divided into 32 exponentially-spaced temperatures: 300, 302, 304, 306.1, 308.2, 310.3, 312.4, 314.5, 316.7, 318.8, 321, 323.2, 325.4, 327.6, 329.8, 332, 334.3, 336.7, 338.8, 341.1, 343.5, 345.8, 348.1, 350.5, 352.9, 355.3, 357.7, 360.1, 362.6, 365, 367.5 and 370. The first temperature of 300 K is lower than the

temperature of our MD simulations (310 K), while the last temperature of 370 K is well above physiological temperatures but below the disruption phase transition of our bilayer, allowing it to keep its shape. We use one of our previously made initial structures for PACE MD simulations, the bent-half structure, as the initial structure for all 32 of the temperature replicas. In these sets of simulations, each simulation box contains 400 POPC lipids, 1 SP-B protein, 7 Cl^- counter ions to neutralize the system and $\sim$ 8000 water molecules. Prior to commencing REMD, we first equilibrate each of the 32 systems at its own temperature for 2 ns in a simple MD run. In the production REMD step under NPT conditions, using Parrinello-Rahman semi-isotropic pressure coupling with τ_p = 5 ps and compressibility = 4.5×10^{-5} bar^{-1}, we run the all 32 systems in parallel for 3 μs. We use a time step of 4 fs. Exchange attempts between systems with neighbouring temperatures happens every 800 steps.

3.2.2 Other Computational Details

We use GROMACS tools and the STRIDE and salt bridge plugins of VMD 1.9.1 [54, 55] to analyze the simulation results. We use VMD 1.9.1 also to visualize our system. We run all-atom simulations mainly using SCINET GPC clusters [56].

For the all-atom simulations, we typically run the system in parallel on 400 CPU cores. On average, each simulation runs at a rate of 16–20 ns/day. In practice, it takes approximately 400 days to run all six of the all-atom simulations, based on the queue availability of our computational resources. In the case of the PACE simulations, we use 120 cores, yielding $\sim$260 ns/day. It takes almost three weeks to simulate for 3 μs. In the case of REMD simulations, we use 16 cores per replica, for a total of 512 cores,

yielding a simulation rate of 100 ns/day. These simulations take almost 5 weeks to run.

3.3 Results

In our OPLS-AA simulations, there are two simulations in which the protein maintains an open structure just inside the bilayer and parallel to it, two in which a bent protein stabilizes defects within the membrane, and two for which the protein forms a globular structure in the water environment outside the bilayer. In the PACE simulations, all of our systems end up with open structure, parallel to lipid bilayer. In all our simulations, the proteins do not reach and have no interaction with the preformed pore that was included in the initial lipid bilayer to ease the movement of lipid molecules between the bilayer leaflets.

3.3.1 MD Simulations

In our sets of MD simulations, we use both open and bent (V-shape) conformations as starting structures (**Figure 3.4**, **Figure 3.5** and **Figure 3.6**), and employ both OPLS-AA and PACE force fields. A brief overview of each simulation is provided in **Table 3.4**. In the following, we present quantities that give a sense of how the systems are approaching equilibrium, a narrative of how the protein evolves in time and structural analysis pertinent to how the protein interacts with the bilayer.

System	Number of atoms	Time (ns)	Description of protein behaviour
open-in-OPLS-AA	179,069	2,000	Stable structure
open-half-OPLS-AA	179,069	2,000	Stable, but sinks to similar depth as open-in-OPLS-AA
open-out-OPLS-AA	179,069	2,500	Forms closed structure after large structural changes
bent-in-OPLS-AA	179,069	2,500	Relaxes, but keeps overall positioning in membrane
bent-half-OPLS-AA	179,069	2,500	Forms pore; V-shape opens up
bent-out-OPLS-AA	179,069	700	Protein closes rapidly
open-half-PACE	13,072	3,000	Protein remains in the water lipid interface
bent-half-PACE	17,029	3,000	Protein opens up, pushes lipid head groups to place itself in the water lipid interface

Table 3.4: Brief summery and descriptions for each simulation.

3.3.1.1 System in Equilibrium

In the case of the all-atom simulations, despite the long run times, we see from the plots of the potential energy (**Figure 3.7**) that the systems are still relaxing. However, from the nearly steady-state behaviour of the radius of gyration (Rg) of the protein's backbone (**Figure 3.8**) and the small root mean square deviation (RMSD) values attained at the longer times (**Figure 3.9**), we have some indication that each of the simulations is converging to an equilibrium state. In contrast, the PACE simulations appear to equilibrate after approximately 1 μs according to the potential energy time and RMSD time series, even though Rg still exhibits some evolution; thus, system energy may not be an optimal quantity for gauging equilibration for the PACE simulations.

We note that for the all-atom simulations, the energy plots are very close to each other because of the identical contents of the simulation boxes. For the systems using

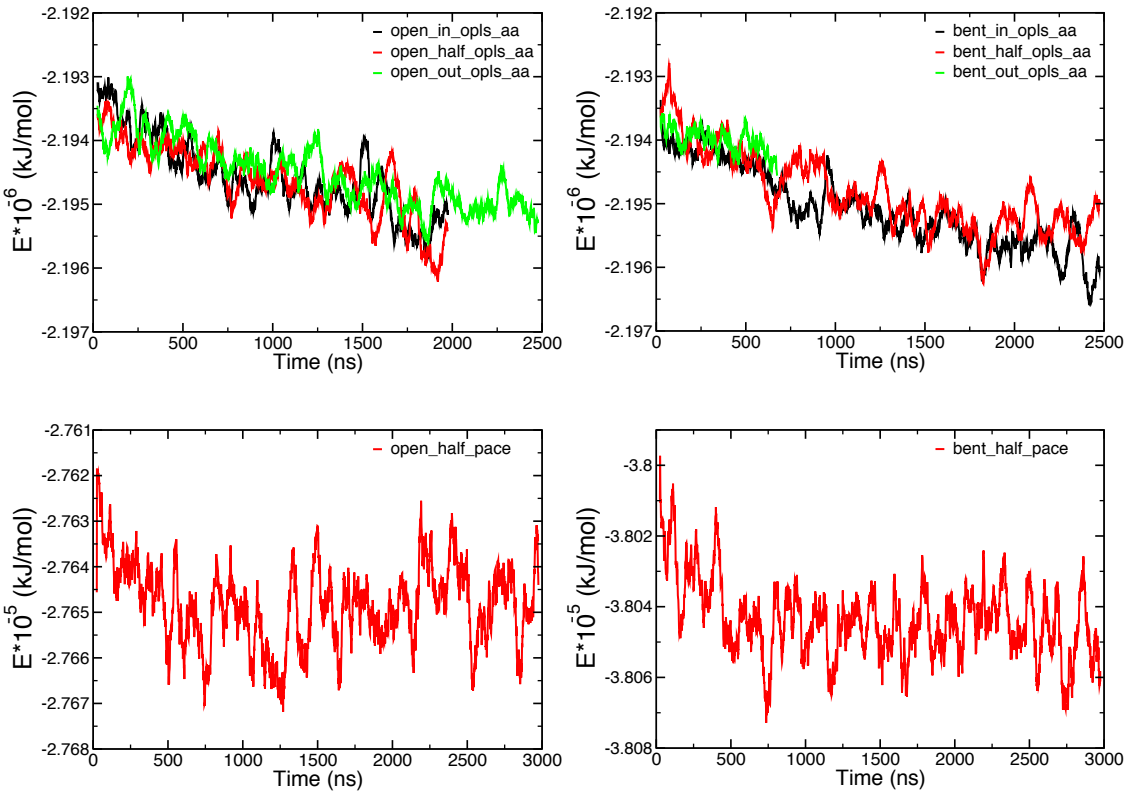

Figure 3.7: Time evolution of the potential energy of open (left) and bent (right) systems in OPLS all-atom simulations (top) and PACE (bottom) during representative simulations. In these panels, shown are running averages over 50 data points, with 1 ns between data points.

PACE force field simulations, as the number of lipid and water molecules are different in each simulation, the values of the potential energies are significantly offset from each other.

We use radius of gyration plots to study the shape of our proteins qualitatively during the simulation and to find geometrical similarities in our proteins, e.g. open structures have similar Rg values (**Figure 3.8**). We calculate the Rg of the protein's main chain, during the simulation for both all-atom and hybrid-model simulations. The initial radius of gyration in OPLS-AA simulations for bent and open are ∼1.6

nm and ∼2.1 nm respectively. Small differences in initial values are due to small changes in protein structure that arise during preparation of the simulation box. In general, the final Rg values for each simulation correlate nicely to the final snapshots of the protein in **Figures 3.4**, **Figures 3.5** and **Figures 3.6**, although the similarity in Rg between the bent-half and the open simulations illustrates the limits of relying on Rg to differentiate between structures i. e. the bent-half structure has similar Rg with open-in and open half, but its tertiary structure is quite different from these two proteins.

For the case of the bent structures, the proteins in each different system reach different Rg values, which represent different final protein structures in these sets of simulations (**Figure 3.8**). In the bent-out structure, the protein closes into a compact structure in the first steps of simulation, which leads to the big drop in Rg during the simulation. For both the bent-half and the bent-in structures, the protein starts with similar structures but the structures diverge as the simulations continue.

In the open-out simulation, the protein structure closes after dramatic changes in the Rg value. These changes arise as the protein's structure optimizes in the presence of water environment. In the open-half and open-in systems, despite the small fluctuations, the Rg time series have similar trends and reach similar values.

Similar to the all-atom simulations, we calculate the Rg of the protein's main chain for hybrid-model force field simulations. We can see that in the open-half-PACE system, apart from some notable dips, Rg remains at roughly its initial value (**Figure 3.8**). In bent-half-PACE, the protein opens up slightly to produce a more extended structure compared to its initial structure (**Figure 3.8**).

To further investigate conformational changes of the protein's structure, we cal-

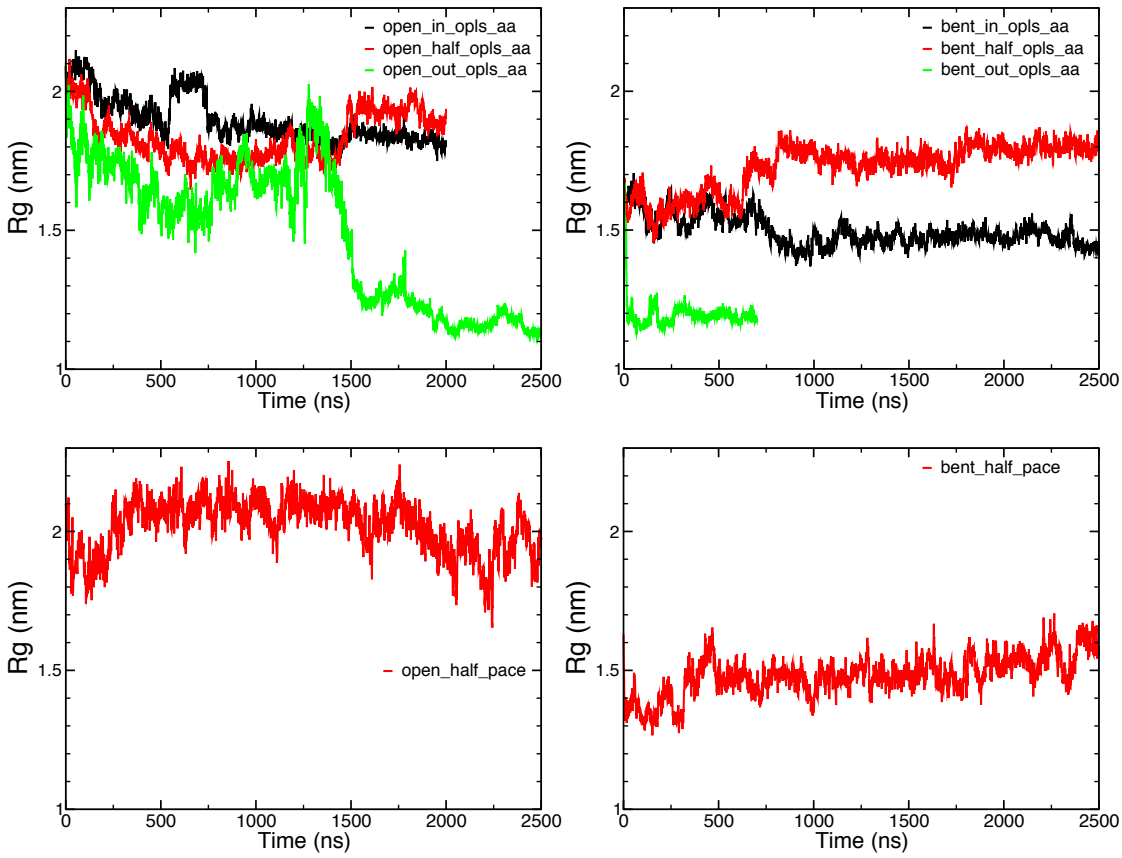

Figure 3.8: Time evolution of the radius of gyration of open (left) and bent (right) systems in OPLS all-atom simulations (top) and PACE (bottom), during representative simulations. We have used the heavy atoms of the main chain of the protein to calculate the Rg, with 1 ns between data points.

culate RMSD values for the main chain of the protein for each system with respect to the final configuration of each trajectory of the system for both all-atom and hybrid-model force field simulations (**Figure 3.9**). For the all-atom simulations, the RMSD plots show that each system has reached an RMSD value of less than 2 Å, a value often taken to indicate that the system has reached an equilibrium state (**Figure 3.9**) [60, 61]. In all of the simulations we can see the RMSD value is decreasing roughly monotonically, except in the open-out system, where the RMSD value decreases initially, but then increases as the protein significantly reconfigures itself. The

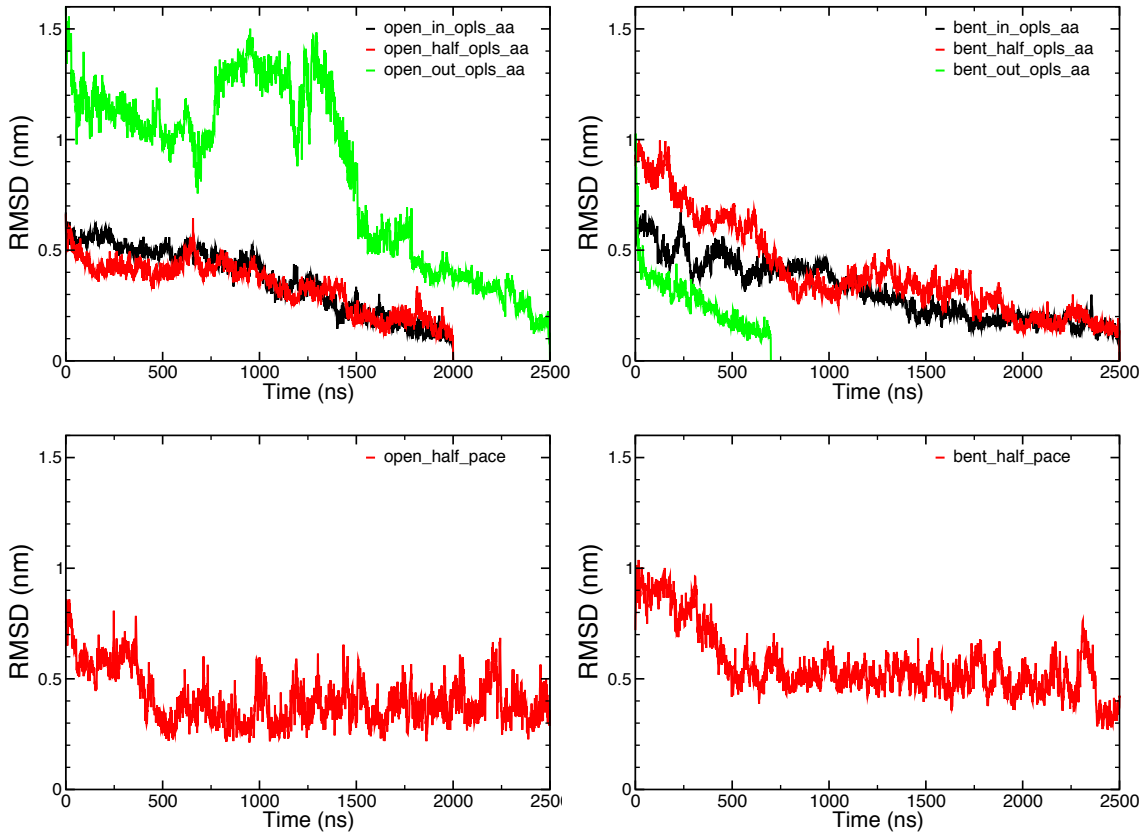

Figure 3.9: Time evolution of RMSD (both rotational and translational) of open (left) and bent (right) of the system in OPLS all-atom simulations (top) and PACE (bottom) during representative simulations compared to the final structure of the protein. We have used main chain of the protein to calculate the RMSD, with 1 ns between data points.

RMSD dramatically decreases after that, indicating that the protein is settling into an equilibrium conformation (**Figure 3.9** and **Figure 3.12**).

In hybrid-model force field simulations, similar to all-atom simulations, the RMSD values approach 2 Å by the end of the simulation (**Figure 3.9**). Our results indicate that the PACE simulations have larger fluctuations in RMSD, as compared to OPLS simulations.

3.3.1.2 System Evolution During the Simulation

In **Figures 3.10** to **3.17**, we can see snapshots and plots of how each system evolves during the simulation. In all of these simulations, the protein evolves in such a way that the charged residues face the lipid head groups and water molecules.

In the open-in-OPLS-AA simulation, the position of the protein remains fairly stable during the simulation (**Figure 3.10**). The protein is initially placed in a lipid chain environment with hydrophilic residues facing toward lipid headgroups and water molecules. The secondary structure of the protein changes during the simulation but the protein retains its overall position in the lipid bilayer. The loop between the central two alpha helices of the protein, is mobile amongst the lipid chains, and interacts with the lipid chains of the opposite leaflet (t= 640 ns) but not to the extent of perturbing it.

In the open-half-OPLS-AA simulation, the protein sinks deeper into the lipid bilayer, reaching a position similar to the open-in-OPLS-AA system (**Figure 3.11**). In this simulation, the protein is also initially placed among the lipid head groups with hydrophilic residues facing toward water molecules. Here we can see the loop between the central two alpha helices of the protein maintains more stable among lipid headgroup in comparison to the open-in-OPLS-AA simulation just described above.

In the open-out-OPLS-AA simulation, the protein undergoes rather dramatic changes in its structure during the simulation. In this simulation we can see the flexibility and changes in the orientation and structure of the the two middle helices of the protein. The loop between the central two alpha helices also shows its flexibility,

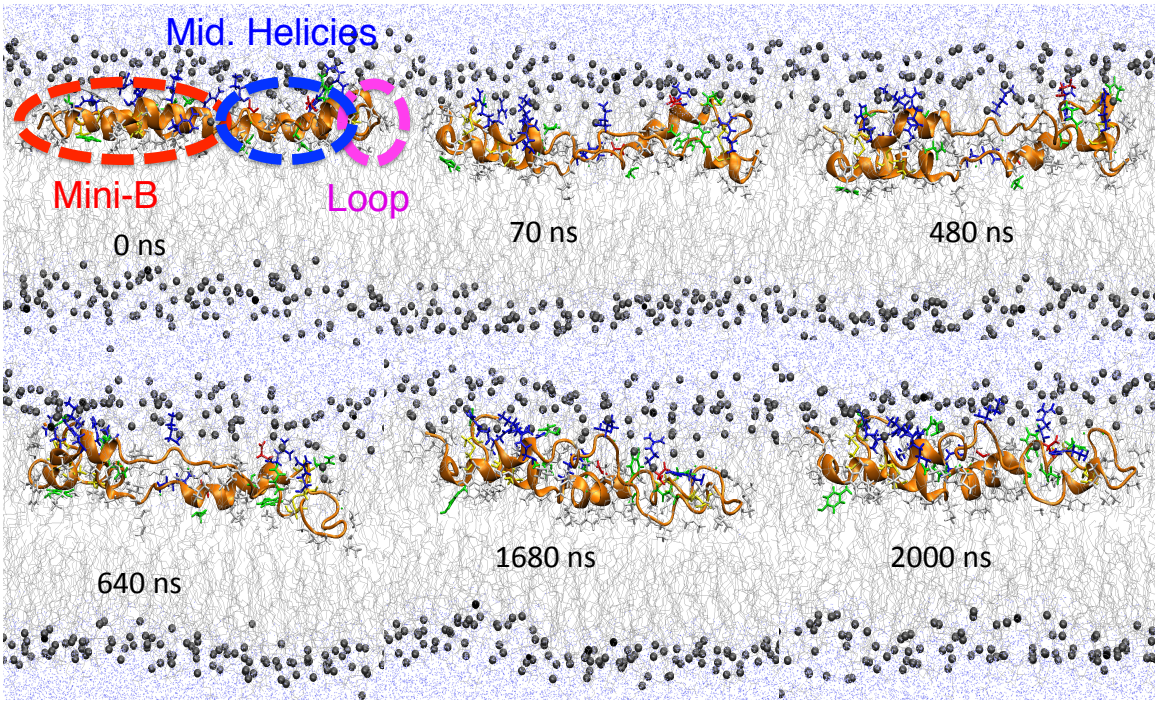

Figure 3.10: Snapshots of open-in-OPLS-AA system during the simulation. We can see protein initially has an open structure, and is positioned just at the polar/apolar interface of the. Hydrophobic sidechains (white coloured) are facing towards the centre of bilayer, while the hydrophilic sidechains (coloured) face toward the water molecules (blue dots). As the simulation progresses the N- and C-terminal α-helices, i.e. the Mini-B region, (the helix pair on the left) largely retain their helical structure, while the "middle" helix pair (the helix pair on the right) show larger changes in secondary structure. The loop between the middle helix pair (as shown in the far right of each snapshot) dips, but does not reach the other lipid leaflet.

which seems to help the protein reach to its final structure. It starts from an initially unfavourable orientation, where the hydrophobic residues face towards the charged headgroups of the lipid molecules. During the simulation, the secondary structure of the protein changes in a way that the protein forms a closed structure with hydrophobic residues prevented from contacting water molecules. The charged groups in the Mini-B part of the protein make the initial contacts with lipid headgroups and these

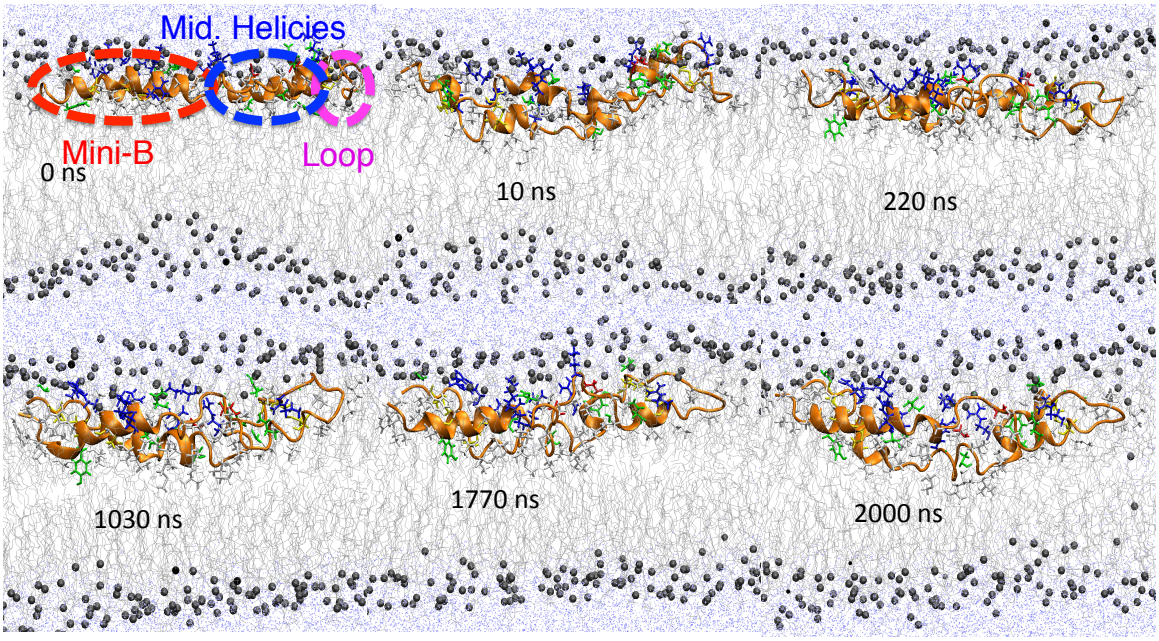

Figure 3.11: Snapshots of open-half-OPLS-AA system during the simulation. We can see protein initially has an open structure, placed among the lipid headgroups half-inside the lipid bilayer. Hydrophobic sidechains (white coloured) face towards the centre of bilayer, while the hydrophilic sidechains (coloured) face toward the water molecules (blue dots). We can see that during the simulation the protein sinks deeper into the lipid bilayer and the secondary structure of some regions of the protein changes dramatically.

contacts remain during the simulation (**Figure 3.12**).

In the bent-in-OPLS-AA simulation, the protein remains inside the lipid bilayer, preserving the water-containing defects initially produced in the bilayer during the embedding of the protein (**Figure 3.13**). The initial sharply V-shaped structure of the protein changes during the simulation, to form a more contiguous curve.

The most interesting system of these eight simulations is bent-half-OPLS-AA. In this simulation, The loop between the central two alpha helices of the protein, via Pro-39 residue, reaches across to the other leaflet and penetrates the membrane, producing a pore in the lipid bilayer (**Figure 3.14**). Proline, based on Wimley-

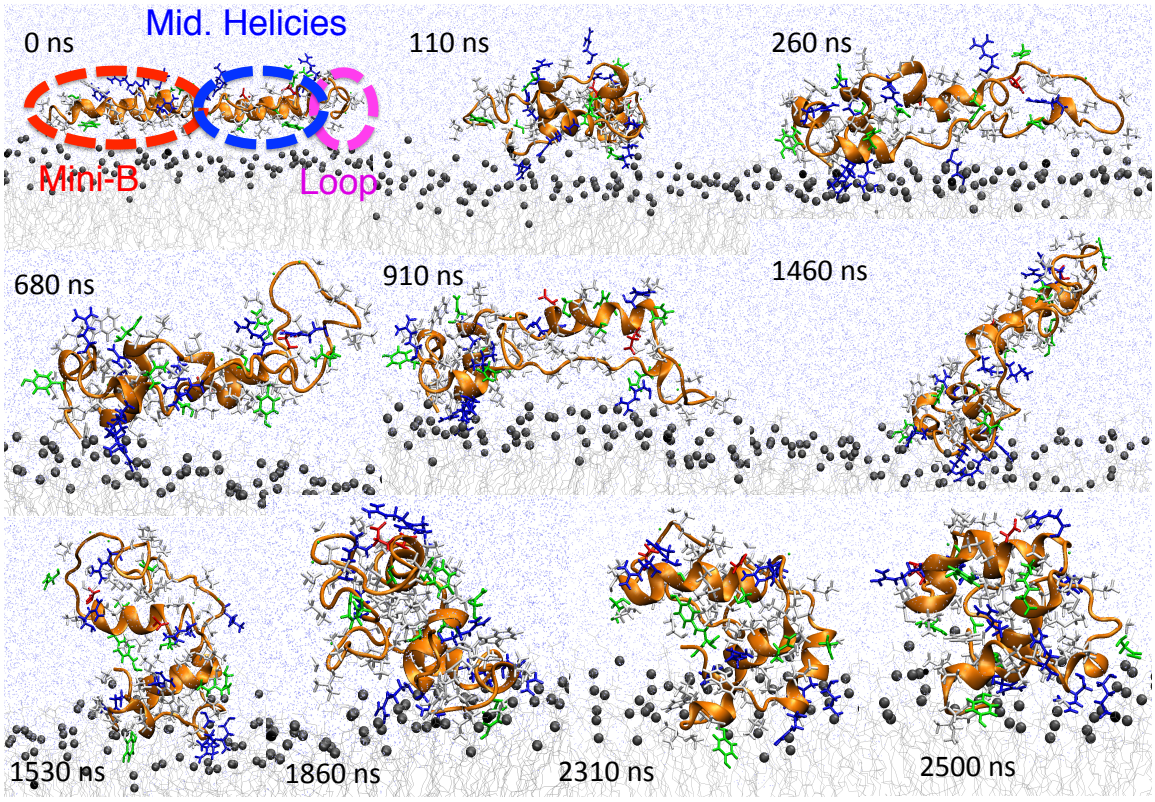

Figure 3.12: Snapshots of open-out-OPLS-AA system during the simulation. We can see protein initially has an open structure, placed outside the lipid bilayer. Hydrophobic sidechains (white coloured) initially face toward the lipid bilayer, while the hydrophilic sidechains (coloured) face toward the water molecules (blue dots). We can see during the simulation the protein's structure changes dramatically to take on a closed structure.

White whole-residue hydrophobicity scales [62], is a hydrophilic amino acid that in this simulation reaches to the headgroups of the other bilayer leaflet, penetrates the bilayer. **Figure 3.15** shows changes in the position of the Pro-39, with respect to the lipid bilayer during the simulation. Here, in addition to the average position of the phosphorus atoms within the lipid headgroups of each leaflet, we show the positions of the phosphorus atoms located within a cylinder of radius 2 nm (parallel to the bilayer normal) around the Cδ carbon atom of the Pro-39 residue, as well as the

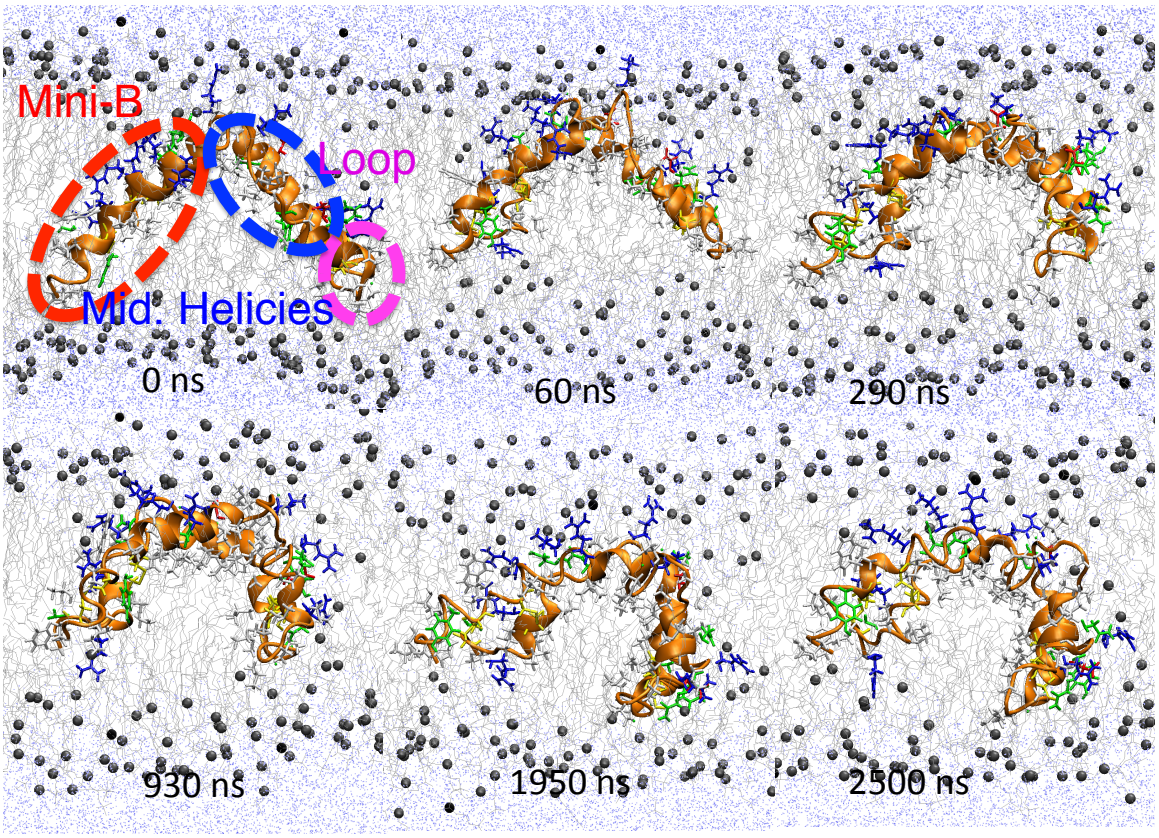

Figure 3.13: Snapshots of bent-in-OPLS-AA system during the simulation. We can see protein initially has a bent structure, placed inside the lipid bilayer. Hydrophobic sidechains (white coloured) face towards the centre of bilayer, while the hydrophilic sidechains (coloured) face towards the water molecules (blue dots) and lipid head groups. There are water-containing defects initially produced in the bilayer during the preparation steps of this system, which we can see them persevered during the course of the simulation. We can see the initial sharp bend structure converts into ca continues curve, located inside the bilayer.

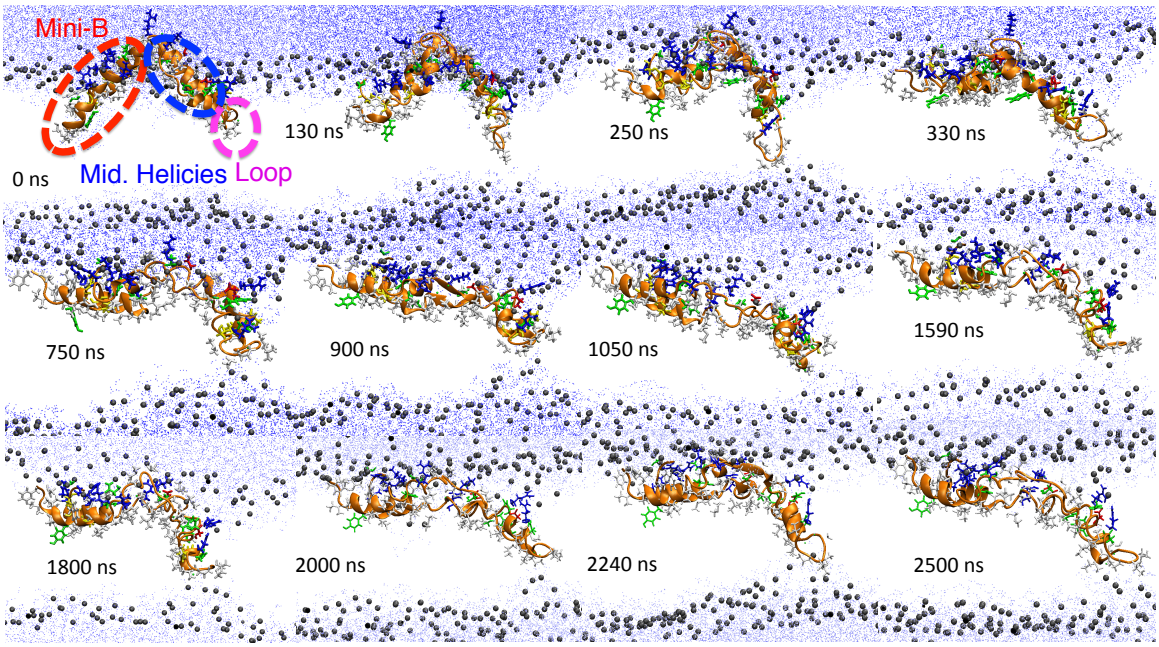

Figure 3.14: Snapshots of bent-half-OPLS-AA system during the simulation. We can see protein initially has a bent structure, placed half inside the lipid bilayer. Hydrophobic sidechains (white coloured) face towards the centre of bilayer, while the hydrophilic sidechains (coloured) face towards the water molecules. We can see during the simulation that the Mini-B part of the protein moves toward the lipid-water interface, parallel to the lipid bilayer, while the middle part of protein (middle helices and the loop between them) reaches the other leaflet and produces a pore-like defect in the bilayer. During the simulation, we can see changes to the secondary (β-strands at t=900 ns) and tertiary structure of the protein.

position of that carbon atom itself. This carbon atom tracks the action of the proline residue within the loop region as it takes part in forming the pore, and the time series shows instances of dipping down towards the bottom leaflet before progressing continuously closer to the bottom leaflet, stabilizing the pore formed, aided by Pro-39 (**Figure 3.14** and **Figure 3.15**).

In **Figure 3.16** we can see the changes in the order parameter of the lipid acyl-

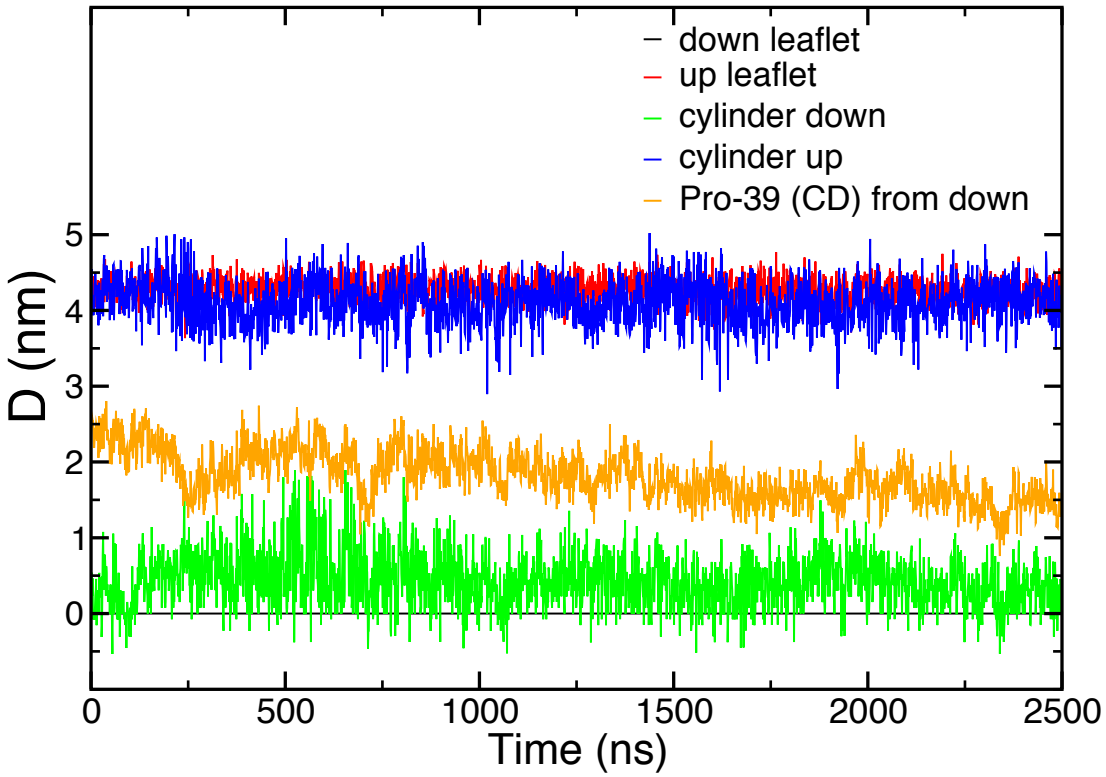

Figure 3.15: Position of Pro-39 (orange) with respect to the lipid bilayer. The black and red colours represent the position of bottom and top leaflets of the membrane, respectively. The blue and green colours represent the position of phosphorus atoms of the lipid bilayer near to the Pro-39 residue. We can see the protein first touches the lipid headgroups of the bottom leaflet at about 250 ns and remains near to the leaflet for the rest of the simulation.

chains with respect to the bilayer normal for the lipids that have the phosphorus atom of their headgroup located within a cylinder of radius 2 nm (parallel to the bilayer normal) around the Cδ carbon atom of the Pro-39 residue. Although there is no direct point-to-point relation between the changes in the depth of the loop part of the protein (**Figure 3.15**) and the order parameter of the acyl-chains (**Figure 3.16**), we can see that during the penetration of the lipid bilayer ($0 - \sim 1$ μs), the order parameter of the acyl-chains displays greater fluctuations. On the other hand, when

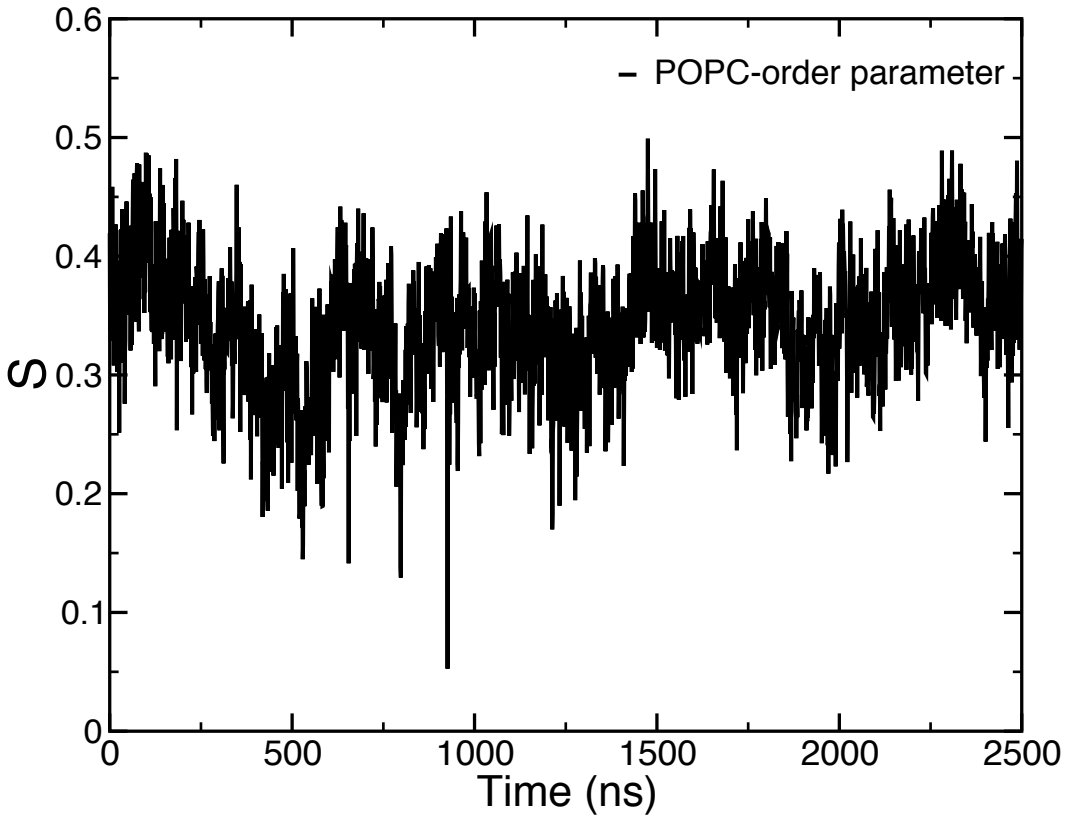

Figure 3.16: Average order parameter of lipid acyl-chain in bent-half-OPLS-AA system during the simulation with respect to the Z direction. The calculations are done for lipid molecules with their head group within a 2 nm radius of a cylinder parallel to bilayer normal with Pro-39 in the centre.

the protein is in a stable position in the lipid bilayer, the fluctuations in the order parameter decreases.

In the bent-out-OPLS-AA system, the protein makes a closed structure in the first few nano seconds of the simulation, remaining closed for the remainder (**Figure 3.17**). Here, unlike the open-out-OPLS-AA system, the Mini-B part faces toward the water molecules, while the charged residues in the loop between two middle helices interact with the lipid headgroups.

Visual comparison of the final structures of the protein in open-in-OPLS-AA and

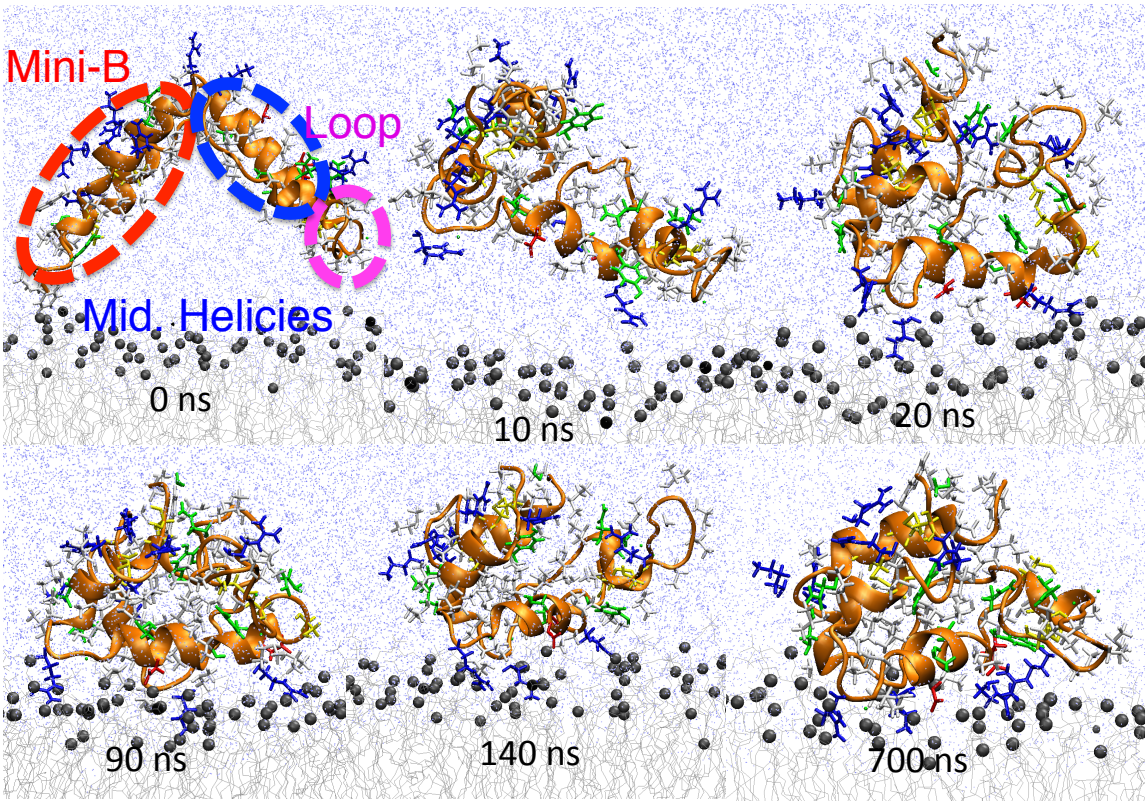

Figure 3.17: Snapshots of the bent-out-OPLS-AA system during the simulation. We can see that the protein initially has bent structure, placed outside the lipid bilayer. Hydrophobic sidechains (white coloured) face towards the centre of bilayer, while the hydrophilic sidechains (coloured) face towards the water molecules. We can see that SP-B forms a closed structure very soon after the start of the simulation.

open-half-OPLS-AA simulations (**Figure 3.18**) and similarities in residue-residue distances plots (**Figure 3.19**) indicate at least partial convergence in these two structures. The RMSD value for the last 50 ns of the open-in simulation compared to the last structure of the open-half system is 4.75±0.13. Similarly, the RMSD value for the last 50 ns of the open-half simulation compared to the last structure of the open-in system is 4.95±0.14. However, for the last 50 ns of bent-half system, using the last structure of open-in and open-half simulations as the reference structures provides higher RMSD values of 5.43±0.29 and 7.353±0.29, respectively. In addition to visual

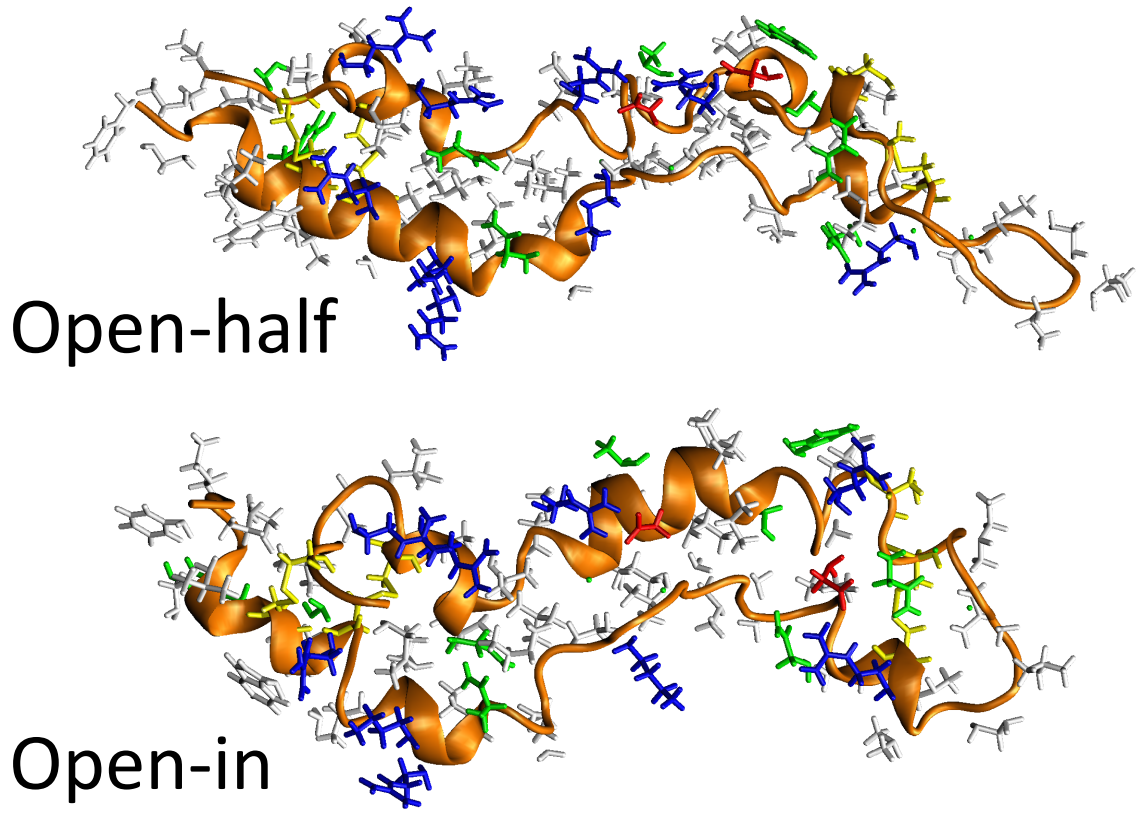

Figure 3.18: Comparison of the final structures in the open-in-OPLS-AA and open-half-OPLS-AA simulations. We can see similar helical regions in both systems.

comparison and RMSD calculations, in **Figure 3.19**, we can see similar patterns in the close distances for the open-in, open-half and bent-half simulations. Although the bent-half system has completely different orientation in lipid bilayer compared to open-in and open-half systems, the extended structure of bent-half (toward the other leaflet) and the spacing between the residues makes their plots similar. On the other hand, in open-out-OPLS-AA and bent-out-OPLS-AA simulations we can see that although they both take on a closed globular structure, the secondary and the tertiary structures are completely different.

Turning to the PACE force field simulations, in the open-half-PACE system, similar to the open-half-OPLS-AA system, the protein slightly sinks into the lipid bilayer

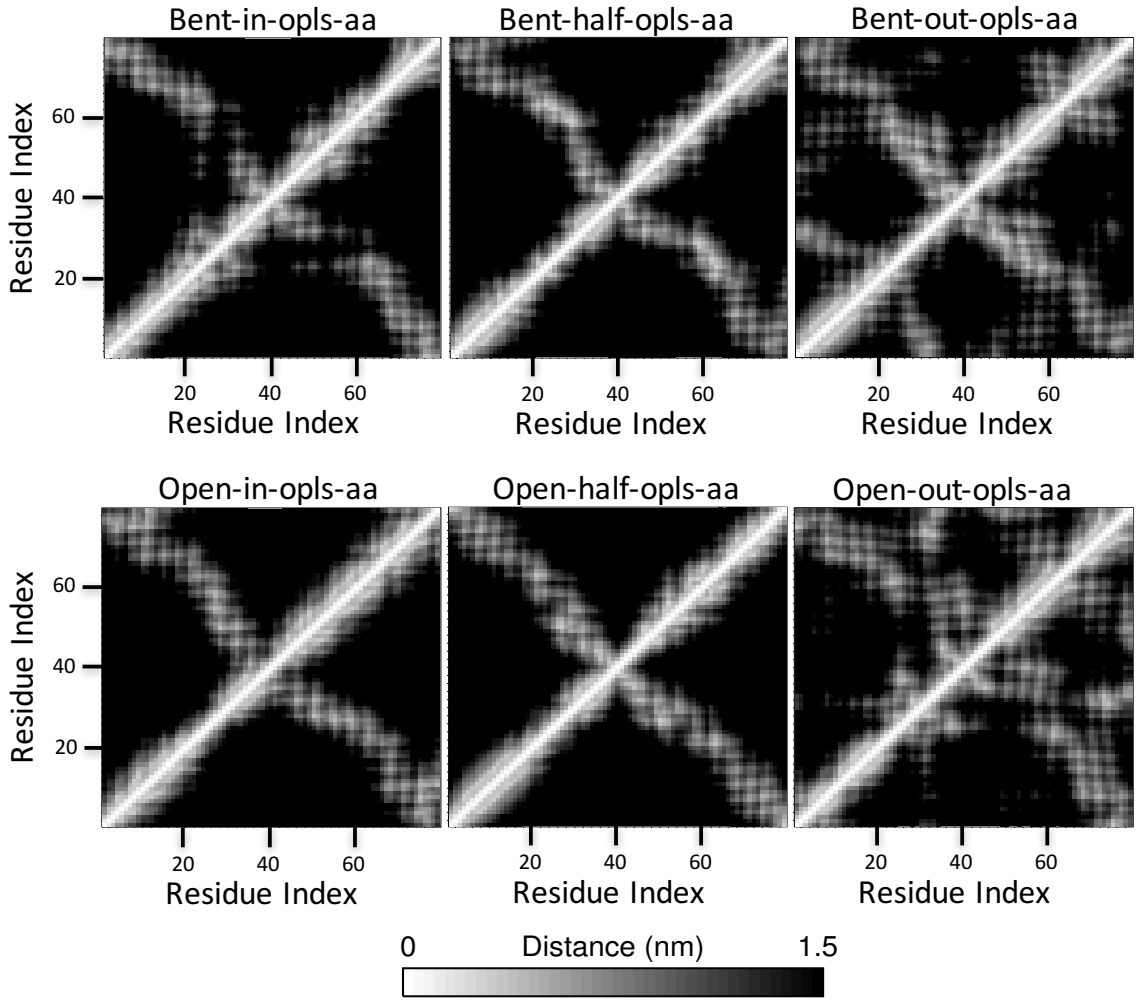

Figure 3.19: Average residues to residues distance over the last 200 ns of simulation. Here, we can see that open-in-OPLS-AA, open-half-OPLS-AA and bent-half-OPLS-AA systems have similar distance plots, indicating similar overall protein structures in these simulations. On the other hand, in both open-out-OPLS-AA and bent-out-OPLS-AA cases the protein forms compact, closed structures.

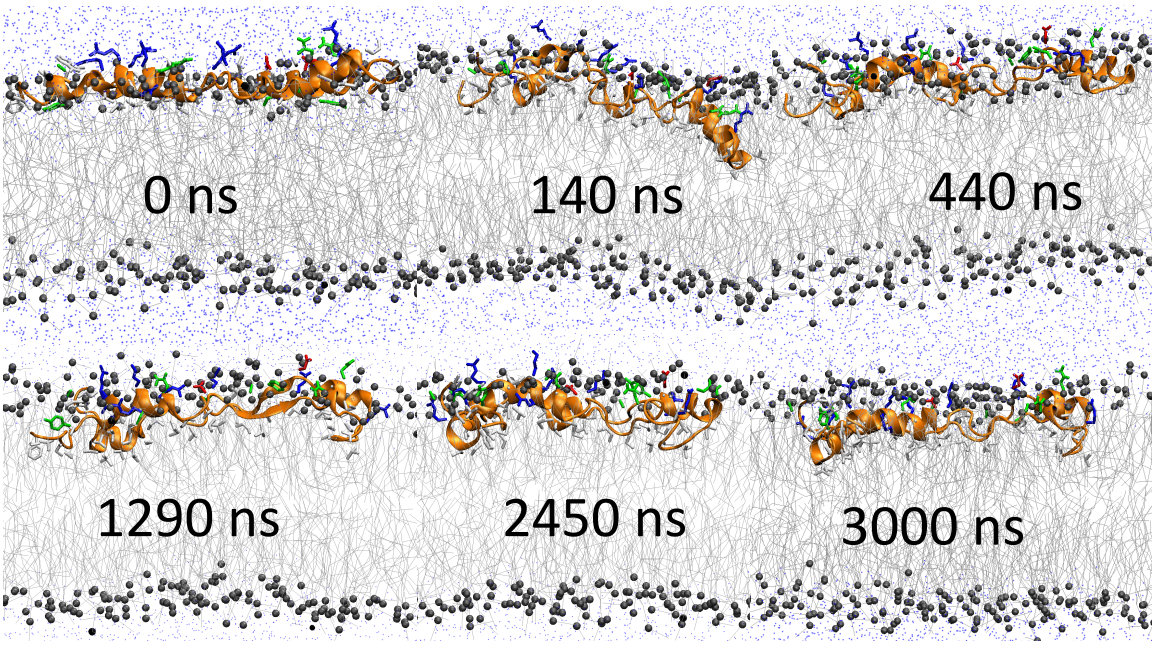

Figure 3.20: Snapshots of the open-half-PACE system during the simulation. We can see that the protein initially has an open structure, placed among the lipid headgroups half-inside the lipid bilayer. Hydrophobic sidechains (white coloured) face towards the centre of bilayer, while the hydrophilic sidechains (coloured) face towards the water molecules (blue dots). We can see during the simulation the protein sinks deeper into the lipid bilayer and the secondary structure of some parts of the protein change dramatically. During the simulation, we can see the loop part dips in but does not reach to the other leaflet.

with the loop part of the protein fluctuating among the acyl-chains of the bilayer (**Figure 3.20**). We can see the loop part of the protein is mobile during the simulation, reaching to the other leaflet of the bilayer, but does not disrupt the bilayer.

In the bent-half-PACE simulation, the protein pushes the lipid headgroups aside to make space for itself to take on an open position at the lipid-water interface (**Figure 3.21**).

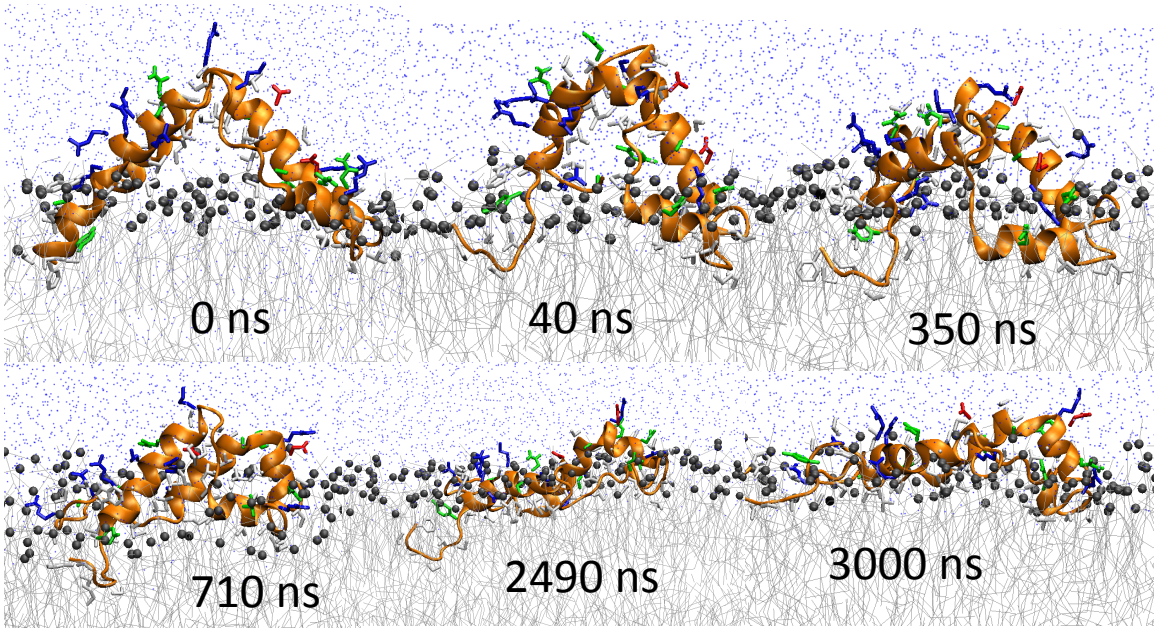

Figure 3.21: Snapshots of the bent-half-PACE system during the simulation. We can see that the protein initially has an bent structure, placed among the lipid headgroups half-inside the lipid bilayer. Hydrophobic sidechains (white coloured) face towards the centre of bilayer, while the hydrophilic sidechains (coloured) face towards the water molecules (blue dots). We can see during the simulation the protein pumps out the lipid bilayer and pushes the head groups to place locate in an open structure, parallel to the water-lipid interface.

3.3.1.3 Protein's Secondary Structure

In addition to the energy, Rg and RMSD plots, we calculate, for a given protein conformation, whether or not each residue is helical. In this way, we determine the percentage helicity (both α-helical and 3-10 helical regions) of the protein as a whole as a function of time and of each residue averaged over a period of time. The initial helicity percentage in **Figure 3.22** and **Figure 3.23**, is calculated based on the structure of protein after applying the forcefield in the first step of setting up simulation in GROMACS. The slight differences between OPLS-AA and PACE

simulations are due to the different nature of each force field and their influence in the secondary structure of the protein.

In **Figure 3.22**, we can see the percentage helicity of the protein during the last 200 ns of simulation. For all of our simulations, we can see that the overall helicity of the protein remains almost constant during this period, indicating that the protein is in near to an equilibrium state in terms of the secondary structure. For all-atom simulations, the mean value of helicity is nearly 35 percent, which is comparable to experimental results [26]. We can see the percentage helicity during this last portion of the simulation time is less than the initial helicity percentage of protein, indicating that our initial model is more helical compared to the equilibrated structure (**Figure 3.22**).

Comparing the percentage helicity of the protein in PACE and OPLS-AA simulations shows that PACE simulations, overall, have slightly higher helicity, especially in bent-half-PACE, with respect to OPLS-AA simulations (**Figure 3.22**). This could indicate the PACE force filed induces more helicity in the protein's secondary structure compared to OPLS-AA force field. The small difference in initial helicity percentages values in open and bent systems (black dots), is due to all the pre-production run setup steps. In the bent-half-PACE system, the protein has $\sim$ 45% helicity, while in open-half-PACE system the protein has, $\sim$ 40% helicity over the lat 200 ns of simulation.

In **Figure 3.23**, we can see the overall percentage helicity of SP-B for each residue over the last 200 ns of simulation for both OPLS-AA and PACE force field systems. The initial structure of the protein has 5 helical structures, helix-I, residues 4$-$7, helix-II, residues 10$-\sim$20, helix-III, residues $\sim$31$-$37, helix-IV, residues 46$-\sim$60 and

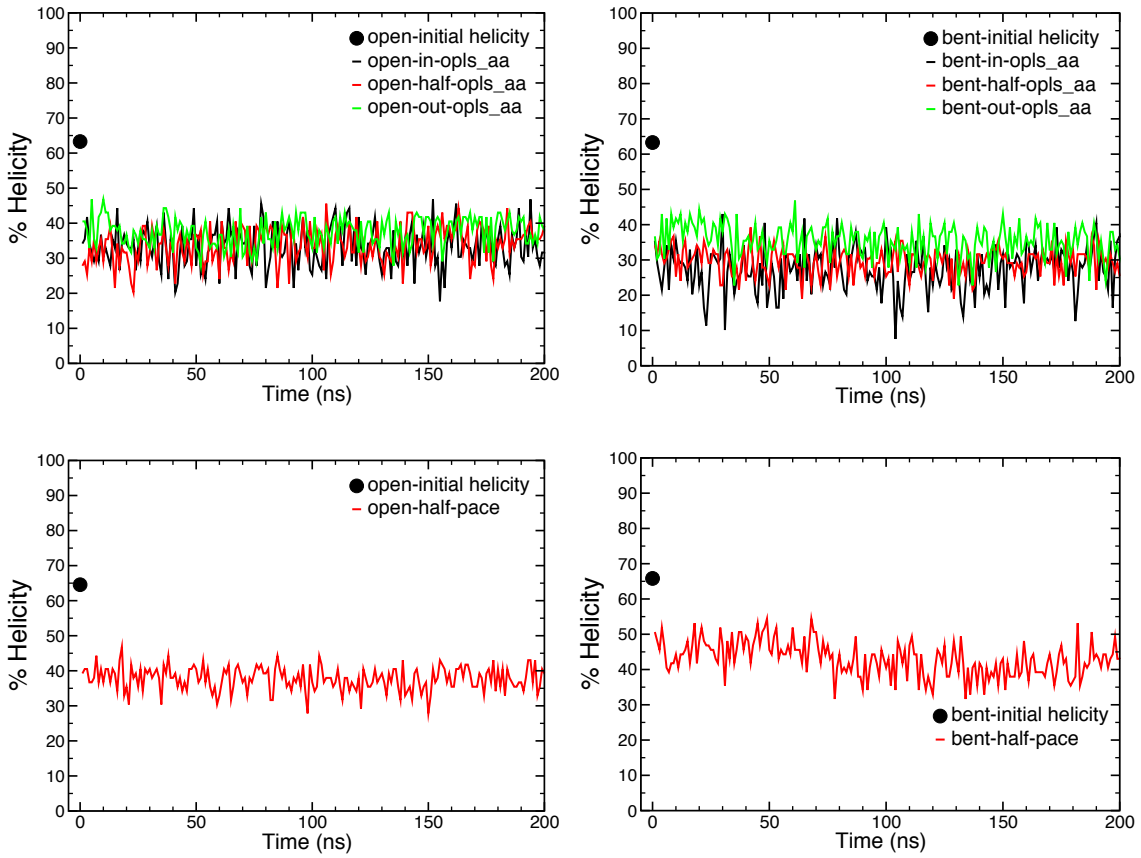

Figure 3.22: Time evolution of percentage helicity of protein in open (left) and bent (right) for initial configuration in OPLS all-atom (top) and PACE (bottom) force fields during the last 200 ns of simulations. We have used the STRIDE plugin of the VMD to calculate the helicity, with 1 ns between data points. The black dot in each plot represents the helicity percentage in initial structure, i.e. before the structure was subjected to all the pre-production run setup steps .

helix-V, residues 67−76 (**Table 3.2** and **Figure 3.23**).

For all-atom simulations, we can see that the helix-I and helix-II regions partially merge into each other, except in open-out system, where the protein has a new clear boundary between helix-I and helix-II. This indicates that the helical structure in the N-terminus should be amalgamated into one helical region. Helix-V, in general, is well preserved in our simulations, which in addition to the helix-I and helix-II regions,

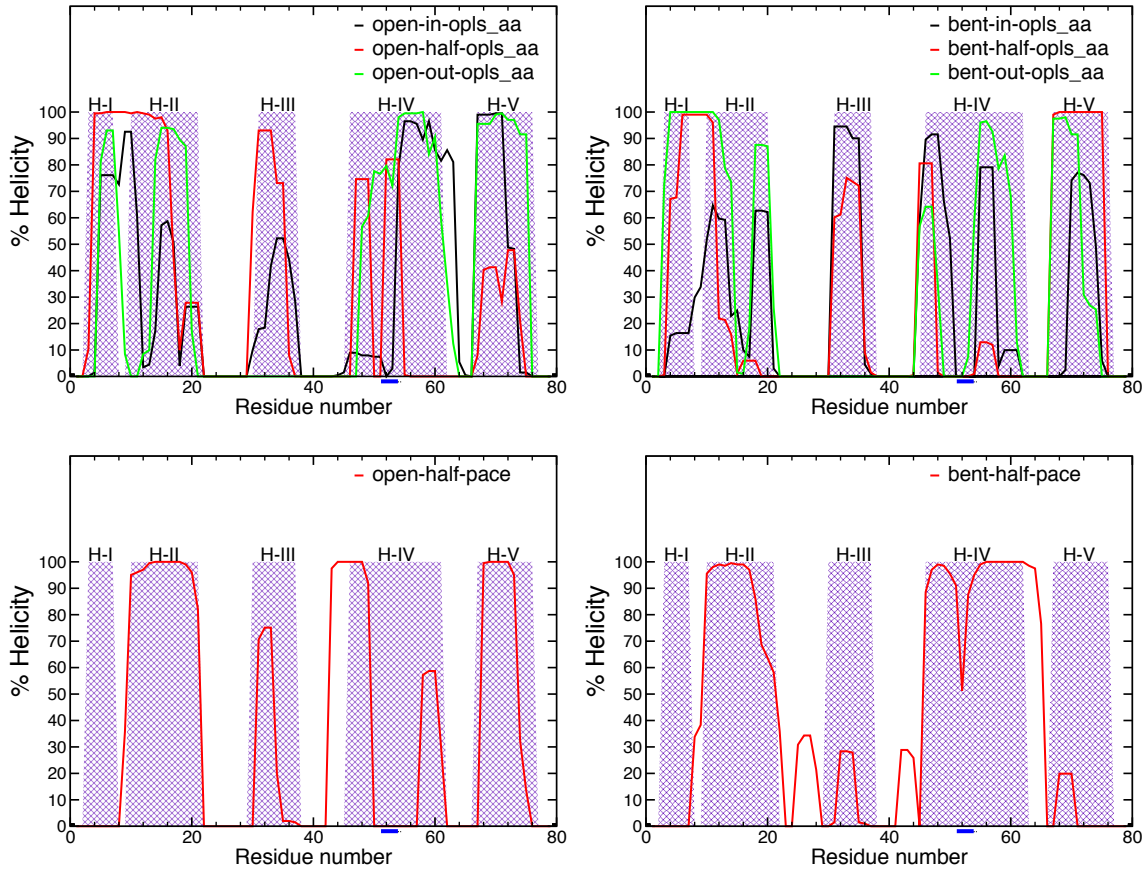

Figure 3.23: Evolution of percentage helicity of the protein in open (left) and bent (right) system in OPLS all-atom (top) and PACE (bottom) simulations for each residue, during the last 200 ns of simulations. We used STRIDE program in VMD to calculate the helicity, with 1 ns between data points. The purple background in each plot represents the helicity percentage of each residue of protein in initial structure for all the simulations, i.e. before the structure was subjected to the pre-production run set-up steps. The blue line on X-axis represents the position for ERYS residues.

indicates that the Mini-B part of our protein confirms the initial helical prediction of our model. Helix-III retains about 60% helicity on average, when the protein is inside or half-inside the lipid bilayer. By contrast the helicity is completely lost, when the protein is in the water environment. This indicates that this part is well predicted by our model, to be helical in lipid environment (open/bent-in/half systems) and non-helical in water environment (open/bent-out systems). In 5 out of 6 simulations, helix-IV divides into two smaller helical regions, except in open-out, where the helicity is more than 70% preserved. When the protein is initially placed inside or half-inside the bilayer, the Glu-51, Arg-52, Tyr-53, Ser-54 (ERYS) sequence of the protein loses its helical structure except in the open-half simulation, which could be due to the presence of salt bridges in this system (discussed in the salt bridges section). This indicates that despite what the homology modelling predicted for the initial structure, residues 51 to 54 are unlikely to take on an alpha helical structure.

For hybrid-model simulations, we can see that the helix-I region loses helical structure, which could be related to its high mobility in the simulation. Our simulations retain more than 90% of the helix-II region. In the bent-half, unlike the open-half, we have new helical regions introduced between the helix-II and the helix-III, as well as the helix-III and the helix-IV. helix-IV is preserved in the bent-half system, while we have a break in helicity in ERYS part in the open-half system. helix-V is well-preserved in open-half and partially preserved in bent-half simulations.

3.3.1.4 Salt Bridges

Salt bridges are electrostatic non-covalent bonds between residues with opposite charges that are sufficiently close to each other [63]. The distance between the two

charged residues, i.e. the oxygen of the negatively charged residue and the nitrogen of the positively charged residue, should be $d \leq 4$ Å to be considered a salt bridge [64]. There are salt bridges present in SP-B's structure, which may play an important role in the protein's secondary and tertiary structure. We used the default value within VMD of $d \leq 3.2$ Å [65] to identify the salt bridges in our system. In **Figure 3.24**, we show the location of salt bridges for the last 200 ns of our OPLS-AA simulations of the protein initially placed inside and half inside the lipid bilayer.

The salt bridges we observe are readily grouped into two categories: a) along the protein sequence and b) across the "arms" of the hairpin, reminiscent of disulphide bonds (the arm is the sequence of SP-B, excluding the loop region). **Figure 3.24** indicates that all of our simulations make at least one salt bridge along the protein sequence i.e. D59-R64. Another salt bridge, from D59-R52 is an interesting one, and only occurs in the open-half simulation. This salt bridge in combination with D59-R64 and E51-R52 form a wall-like structure, in which the side chains stack next to each other (**Figure 3.24**), that prevents the ERYS sequence from losing its helical structure, which it does in the other simulations (**Figure 3.24** and **Figure 3.23**). In the open-half and the bent-half simulations, we have a salt bridge between adjacent amino acids E51-R52. These two residues are thought to play a role in quaternary structure, as they have been proposed to form salt bridges between pairs of SP-B in dimers [42, 44].

The main difference between open and bent simulations is the presence of the D59-K24 salt bridge in the bent structures (both bent-in and bent-half). The location of this salt bridge is quite notable, in that it is located at or near the "hinge" separating the parallel-to-membrane and transverse portions of the protein in the

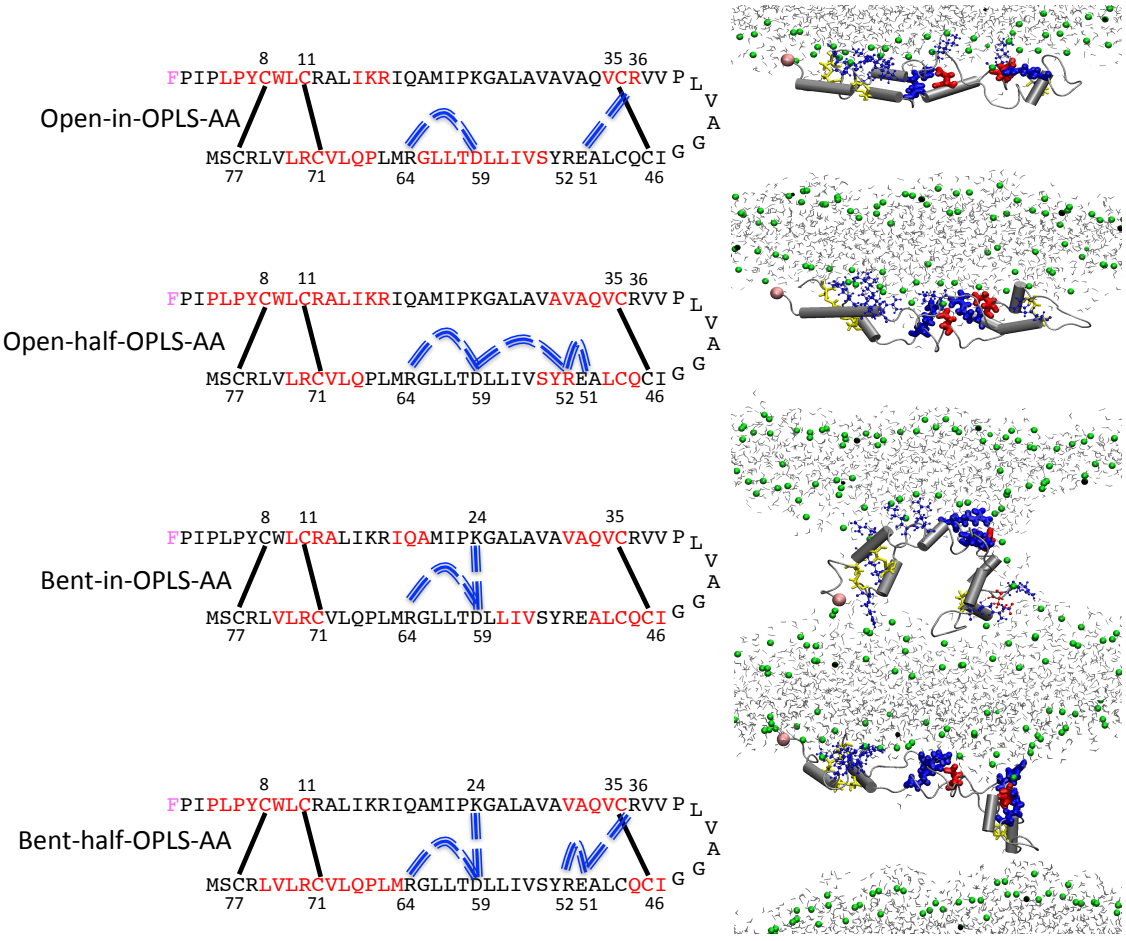

Figure 3.24: Final structures of SP-B in open and bent simulations using OPLS-AA forcefield. On the left, we can see a sequence representation of the protein and on the right, we can see a corresponding snapshot from near the end of the simulation. In the protein sequence, PHE-1 is in pink, the helical regions are in red, the disulphide bonds are in solid black lines and salt bridges are in dashed blue lines. In the snapshots, the PHE-1 is in pink spheres, the protein is in gray, the helical regions are in gray cylinders, negatively and positively charged residues are in red and blue, respectively with the salt-bridge-forming residues bolded, P atoms of lipid head group are in green spheres and water molecules are in silver lines. The last 200 ns of simulation is used for our calculations. 40% presence is chosen as the minimum to show the helicity. We used the default value of d$\leq$ 3.2 Å within VMD to identify the salt bridges in our system.

bent-half simulation. This salt bridge constrains and stabilizes the central portion of the protein (the part between Cys11-Cys71 and Cys35-Cys46 disulphide bonds) and appears to prevent the protein from opening up and widening the "trough" of water molecules between the arms.. We speculate that the twisting of the two arms implicit in the formation of this bridge, as the residue side chains turn from pointing up towards the water to pointing more towards each other, may assist in conformational changes required for bending, i.e., forming or stabilizing the hinge. The other salt bridge across the protein sequence is E51-R36, which is close to the loop as well as the Cys35-Cys46 disulphide bond. The close proximity of this salt bridge to the disulphide bond renders it less important in geometrically constraining the loop portion of the protein.

All of the salt bridges in the bent-half simulation appear after the pore formation (after 600 ns), which could indicate the role of flexibility in pore formation. The D59-K24 salt bridge forms around 600 ns, E51-R52 forms after 1 μs, and both E51-R36 and D59-R64 form after 1.5 μs. In the open simulations, the salt bridges along the protein sequence appear in the early steps of the simulations, except for the D59-R52 salt bridge which appears in the last 800 ns of the simulation. In the bent simulations, these salt bridges are formed after almost 1.5 μs of the simulation. In the bent-in simulation, the D59-K24 salt bridge forms in the first few ns of the simulation. In this system, the E51-R36 and E51-R52 salt bridges appear only in the first few ns of the simulation and disappear in the rest of the simulation. Similarly, in the open-half simulation, the D59-K24 salt bridge appears only in the first few ns and disappears in the rest of the simulation. Thus, we see a breadth of time scales associated with salt bridges. Unfavourable bridges can break up and favourable ones can form, if the

configuration and local environments of the protein are suitable, on the order of ns. By contrast, some salt bridge formation requires on the order of a μs to form, whether along or across the arms of the protein.

The bent-half simulation contains almost all of the possible salt bridges, which is interesting in that this configuration represents a step between an open structure parallel to the bilayer surface (open-in and open-half) and one that is completely bent and disruptive of the membrane (bent-in). Indeed, in **Figure 3.14** we see snapshots for which the protein takes on a long, linear structure, but is tilted by about 30° with respect to the bilayer, e. g. in 1050 ns. We can see in **Figure 3.24** that the N/C terminal regions are parallel to the lipid bilayer, similar to the open simulations, and the central loop interacts with the pore region, similar to the bent-in simulation. Despite the kink between parallel and transverse parts of this protein configuration, the overall structure is similar to that of the open structures, as evidenced by the residue distance plot (**Figure 3.19**). In other words, bending, which results in portions of the protein changing environments in terms of surrounding lipid and water, can occur through localized changes. It is not unreasonable to assume a supportive role played by salt bridges in maintaining protein conformation.

As a salt bridge can form not only between residues, but also between a lipid head group and a residue, one might expect that the propensity to form intrapeptide salt bridges will vary depending on whether the lipids are zwitterionic (as in our POPC bilayer) or anionic (like PG), and this may affect preferred protein structure. In the lung, where a significant portion of the lipid head groups are anionic, the likelihood of forming the cross-arm D59-K24 bridge that appears to stabilize bent configurations, may be different, and hence the bent (membrane disrupting) and open

(parallel to membrane) configurations may occur with different relative frequencies, affecting functionality.

3.3.2 REMD Simulations

Here we present data for REMD simulations, for which each temperature replica started with with the same bent-half-PACE initial configuration. While REMD requires significantly more computational resources as all temperature replicas run in parallel, the presumed advantage stems from the possibility that higher temperatures allow a freer exploration of protein configurations and more fluidity in the lipid bilayer. In our case, while we do recover trends observed in regular MD simulations of bent-half-PACE (**Figure 3.21**), we see a richer set of protein structures in REMD.

In **Figure 3.25**, we show snapshots from the $T = 310.3$ K replica of the REMD ensemble (the T of all of the regular MD simulations). As in regular MD of the PACE potential, broadly speaking, the protein flattens out into an open configuration and remains near the lipid-water interface. We also see (at 565 ns) the protein in a straight, but tilted pose, similar to what was briefly seen in the bent-half-OPLS-AA simulation.

Most strikingly, the first seven (or so) residues of the proline-rich N-terminus make a few excursions through the membrane to the other leaflet (325 ns, 1818 ns, 2662 ns), and otherwise can be seen in a disordered configuration in the region of lipid tails, but not far from the headgroups (595 ns, 2212 ns). These observation supports the idea that this initial portion of the protein serve as an "insertion sequence" [15,16,22]. These excursions of the N-terminus are not seen in the regular MD PACE simulations,

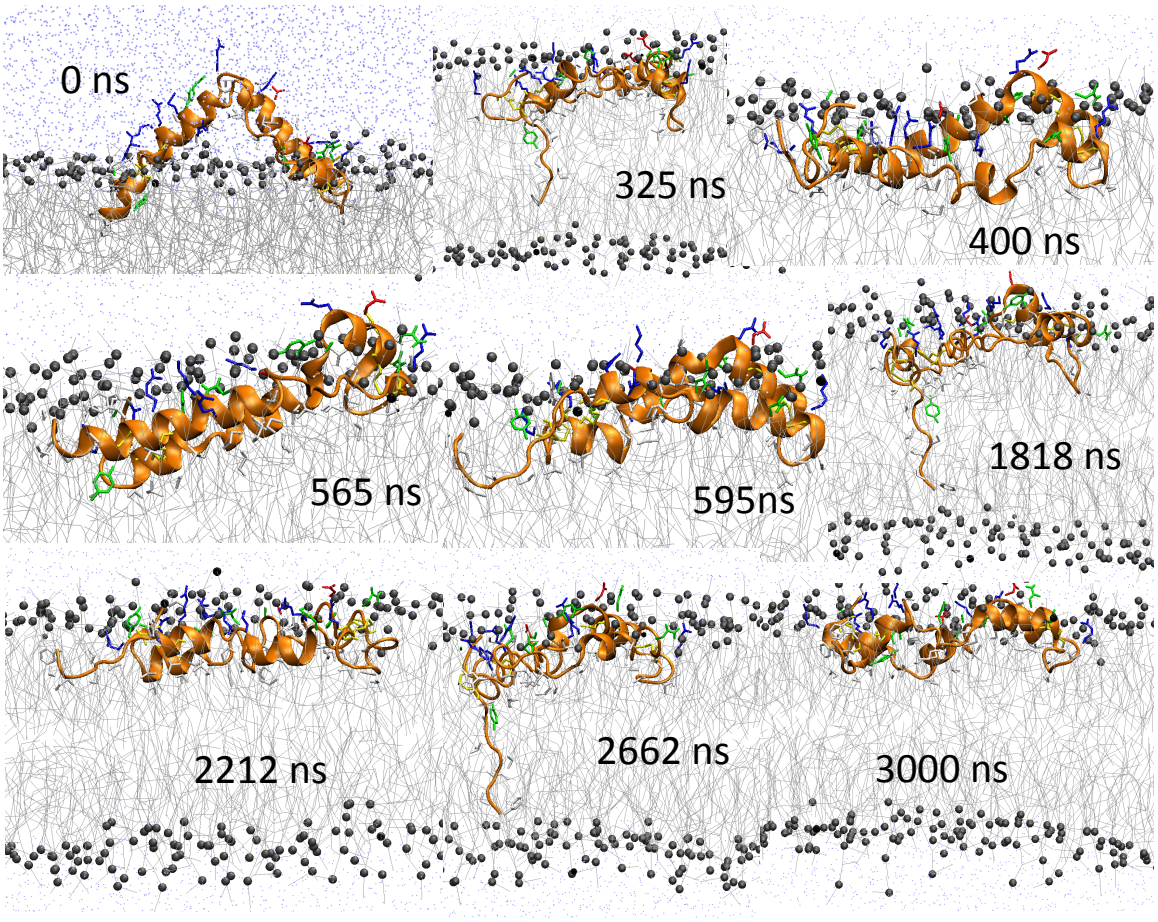

Figure 3.25: Different snapshots of bent-half-REMD system during the simulation at $T = 310.3$ K. We can see that initially SP-B has a bent structure, and is located among the lipid headgroups half-inside the lipid bilayer. Hydrophobic sidechains (white coloured) face towards the centre of bilayer, while the hydrophilic sidechains (coloured) face towards the water molecules (blue dots). We can see during the REMD simulation the protein is in structures, e.g. in 325 ns, 1818 ns and 2662 ns, which were not observed in MD simulations.

and thus we see the benefit of REMD. In contrast to simulations using OPLS-AA, which confers helical-like structure to these first residues, no helical structure is observed for these residues in PACE.

Despite the greater exploration of configurations, we do not see the central loop portion of the protein making excursions deep into the membrane, and this implies that the PACE version of the protein lacks this functionality. The slightly greater overall helicity of the PACE system, particularly in the loop portion (around residue 40), and the lower flexibility this implies, may be responsible for this. Also, the "hinge" portions around D59 and K24 tend to be more helical for PACE than for OPLS-AA. Notably, the D59-K24 salt bridge is all but absent from PACE simulations.

These specific structural discrepancies between PACE and OPLS-AA, namely in helicity of the "insertion sequence" and helicity near the "hinge" and central loop, serve to highlight the connection between flexibility, conferred through lack of helical structure, and the ability to penetrate the membrane.

3.4 Discussion and Conclusions

Obtaining information on equilibrium properties of biomolecular systems from molecular simulation is inherently challenging, given the many orders of magnitude in time separating atomic motions and different conformational changes undertaken by the protein. Our work is based on using all-atom simulations in order to gain some understanding of the basic structural properties of the SP-B protein when interacting with a lipid bilayer. We compare our all-atom results with the coarse-grained PACE potential, which allows longer time scales to be accessed more easily at that expense of

accuracy in describing the physical interactions within the system. Insight from simulation is particularly pertinent given the great difficulty in obtaining experimental data on this extremely hydrophobic protein.

The basic question addressed by this chapter is: what are the energetically reasonable conformations that SP-B can take within or near the bilayer? Our 2-2.5 μs OPLS-AA simulations for the protein within the bilayer express two main answers. First of all, an open structure parallel with and partially submerged within the lipid bilayer appears to be rather stable structure of the protein in lipid-water interface. Second, both "ends" of the protein, i.e., the flexible, largely hydrophobic central two helices and the loop between them, as well as the two more helical termini that remain in close proximity (Mini-B part), are able to retain or even seek out transmembrane orientations. These transmembrane portions can be associated with greater membrane disruption through the stabilization or creation of pores or pore-like structures characterized by the presence of lipid headgroups and water molecules in the membrane interior, and decreases in the value of the order parameter during formation. In all cases, the charged and polar residues maintain proximity to headgroups and water.

When placed outside the bilayer, SP-B rearranges itself into a compact, closed structure, sacrificing helical region H-III to accomplish the folding but increasing helicity elsewhere. While the two closed structures we obtain are different, they both share the characteristic of using one of two hydrophilic "sides" to stick to the membrane, while exposing the other two water. In one case (open-out), H-I, H-II and part of H-V act to anchor the protein to the membrane while hydrophilic residues of H-IV are left away into the water. In the other case (bent-out), H-IV anchors to

the membrane while H-I, H-II and H-V point away. These findings are consistent with coarse-grained MARTINI simulations of Bauokina et al. [27] that show SP-B simultaneously binding to two lipid vesicles while outside the membrane, in that the portion of the closed structure pointing away from the membrane can "accept" and stick to a second, approaching membrane. We also note that the closed structure that forms from the open-out simulation, does so through a rather non-trivial sequence of configurational changes during the course of roughly 1.5 μs.

In adjudicating the degree to which the simulations are equilibrated and hence our confidence in assessing whether the conformations we observe truly reflect stable or metastable protein structures, we note the different time scales on which different structural changes take place. On quite short time scales, less than 10 ns, we see the rather energetically unfavourable initially bent configuration close like a clam outside the membrane (in a mostly aqueous environment). On the order of 100 ns, we observe smoothing out of the initial bend in our V-shaped bent configurations within the bilayer, tipping of the protein that accompanies its bending and straightening while in the membrane (see **Figure 3.14** from 750 ns to 1590 ns) and the bending motion about a hinge (see **Figure 3.14** from 1800 ns to 2400 ns). On the time scale of 1 μs, we see the complete flattening of the initially bent-half-PACE simulation. In this last case, the protein is never in a position to reach across the membrane and all motion takes place near the upper leaflet. Formation or loss of salt bridges also seems to span time scales from 100 ns to 1 μs. With all this in mind, we can reasonably conclude that the open structure situated near the lipid-water interface is a stable, or at least metastable, configuration of SP-B, i.e., apart from optimizing its local interactions and secondary structure, larger scale movements, such as having part of the protein

adopt a transmembrane pose, occur on significantly longer time scales presumably because of a free energy barrier.

With regard to the metastability of configurations in which at least one part of the protein adopts a transmembrane position, we note that the bent-half-OPLS-AA simulation evolves towards the formation of a pore-like defect with the loop portion of the protein progressively engaging with the lower leaflet. The progression towards a configuration strongly implies a free energy minimum associated with the configuration. With regard to the stability of the N/C termini being in a partially transmembrane pose, we can only point to the longevity of the pose, perhaps aided by the dimpling of the bottom leaflet.

Our results regarding salt bridges imply a role for them both in terms of stabilizing or forming the hinge which allows the loop end of the protein to bend towards the bottom leaflet, and in terms of affecting secondary structure as in ERYS segment. The propensity to form intraprotein salt bridges may be affected by the nature of the lipid headgroups, and this implies a role in the local composition of the lipid membrane in the action of SP-B.

We now comment on how well our structures predicted through homology modeling fair for the OPLS-AA and PACE potentials. For the open and bent structures, the helical portions of the protein after lengthy simulation generally fall within the areas predicted by homology modelling. There is a decrease in helicity in the simulated protein, but this only serves to bring the overall helicity of the protein in line with experiments. One notable exception to the agreement between simulation and homology modeling is the polar and charged ERYS sequence in H-IV, which tends to break up the helix. We note that for tertiary structure, our guess at where the protein

bends when adopting a V-shaped configuration is not far off, initially modeled to occur between residues 21-24 and 63-66 but simulation showing a hinge slightly closer to the central loop and located around the salt bridge D59-K24, involving residues 23-25 and 58-60. For the closed structures, we do not make a homology-based prediction, and from a visual comparison, we find little structural similarity between either of them and NK-lysin, confirming the difficulty in predicting SP-B structure from other members of the saposin family.

In comparing the results of the coarse-grained PACE potential from MD and REMD simulations with those obtained with the all-atom OPLS potential, we note differences significant enough to indicate qualitative differences in the way that the protein interacts with the membrane. For PACE, at no point do we observe the central loop portion of the protein (roughly residues 35 to 45) venturing deeply into the bilayer, i.e., probing the opposite leaflet. We do observe, especially within REMD simulations, a variety of open structures with significant differences, e.g., tilted but with a loop portion clinging to the lipid headgroup-water interface, and structures for which the insertion sequence (first 7 residues), but only this sequence, adopts a transmembrane configuration, a configuration that is not supported by experiment [15, 16, 22]. While the overall helicity for both PACE and OPLS-AA is compatible with experiment, the distribution of helicity of PACE is different from that of OPLS. In particular, with respect to OPLS, PACE under-represents helicity in the first seven residues and perhaps also in H-III (residues 30-37), and overemphasizes helicity of residues 41-45, a region directly in the central loop. These observations are consistent with the idea that lack of helicity (implying flexibility) allows portions of the protein to probe the interior of the membrane and high helicity (implying rigidity) makes it

more difficult to do so. Thus, the PACE potential, and therefore the even more coarse-grained MARTINI potential, likely underestimates the ability of SP-B to penetrate, form pores within, and otherwise promote the lipid structures needed for lung surfactant to reduce surface tension.

Keeping in mind the ability, implied by OPLS-AA, of SP-B to be present within the membrane interior, it may be worthwhile to speculatively expand on the detailed picture of the process of membrane fusion obtained from simulations of SP-B and vesicle fusion with the MARTINI potential [23]. This picture could also explain the experimental results of Oelberg, et al [11], on pore formation ability of SP-B in artificial lipid bilayers. In **Figure 3.26**, sequence A-I to A-IV, we present the four paradigmatic stages of membrane fusion: I) contact, II) stalk, III) diaphragm and IV) pore [66], respectively. Sequence B-I to B-IV of the same figure provides schematic representations of SP-B and how it may promote each stage, while sequence C-I to C-IV shows configurations from simulation that provide the motivation for the schematics. For contact (I), the OPLS-AA picture is qualitatively the same as for the MARTINI potential, in that the protein, while at the membrane surface, exposes hydrophilic residues that can "stick" to an incoming membrane and facilitate close contact. General disruption of the membrane by SP-B, especially by configurations tending towards a transmembrane positioning, facilitates both stalk (II) and diaphragm (III) formation. The ability of SP-B to form or to stabilize pores would obviously aid in the final stage (IV) of vesicle fusion.

In summary, 2-2.5 μs simulations using the OPLS-AA potential show that SP-B is stable in an open structure parallel to but slightly embedded in a POPC lipid bilayer. However, the ends of the protein, i.e., both the central loop and the N/C termini, give

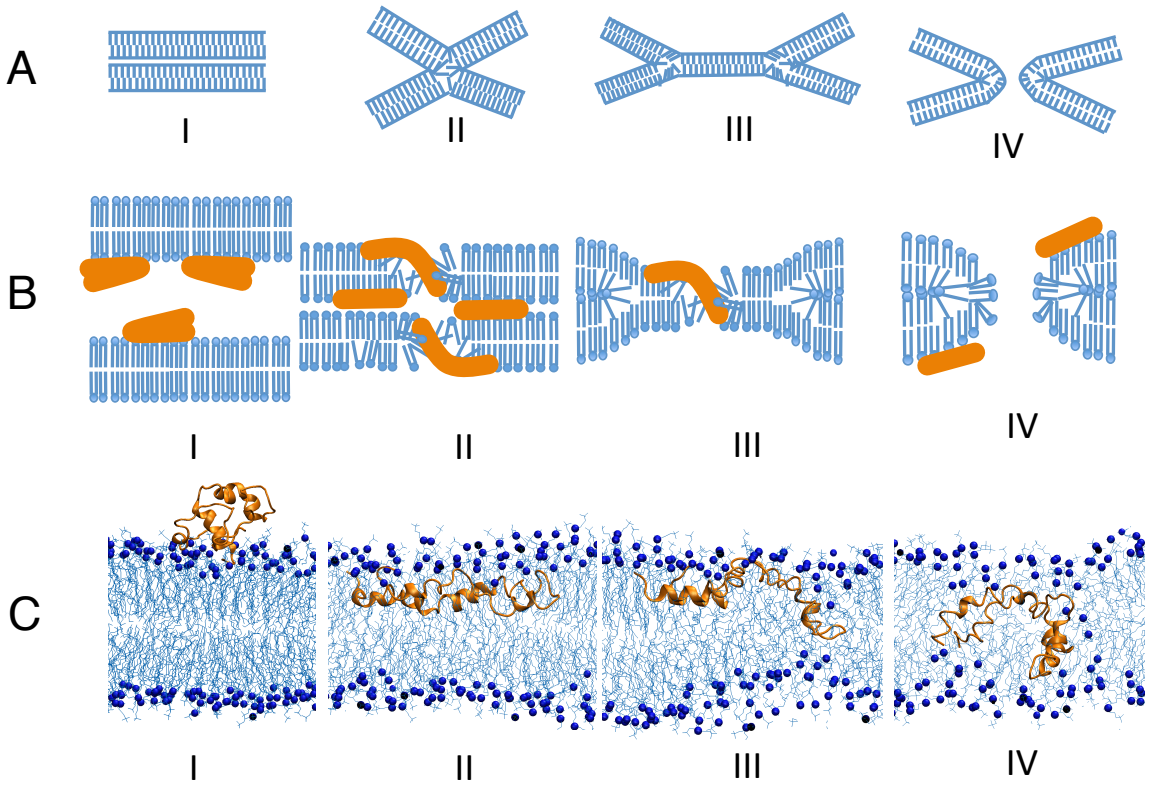

Figure 3.26: SP-B performing fusion of two lipid bilayers. A) the four main steps of fusion: I) contact, II) stalk, III) diaphragm and IV) pore. B) the schematic figures of how SP-B performs the fusion steps: I) anchoring step by SP-B in closed structure, II)lowering energy barrier in bilayer surface to help the bilayer stalk to each other, III) the diaphragm step where SP-B can make pore in lipid bilayer and IV) complete pore. C) snapshots of SP-B simulations: I) outside the lipid bilayer helps to anchor the bilayers, II) in the open structure reduces the surface tension energy barrier, III and IV) performs and maintains the pore.

some indication of their ability to explore the interior of the membrane. In particular, we witness the (largely hydrophobic) loop's ability to reach the opposite leaflet and to stabilize a pore-like structure, aided by the proline residue in it. Lack of strict helicity in certain parts of the protein seems to be a factor in encouraging the adoption of transmembrane positioning, while salt bridges can stabilize bent configurations. Outside the membrane, SP-B is stable in a closed structure. Using the PACE potential results in underemphasis and overemphasis of helicity in different portions of SP-B, and this seems to alter the ability of SP-B to explore the membrane interior, which in turn may affect conclusions about how SP-B assists in membrane fusion.

Bibliography

[1] Postle, A. D., Heeley, E. L. and Wilton, D. C. "A comparison of the molecular species compositions of mammalian lung surfactant phospholipids," (2001) Comp Biochem Physiol, Part A Mol Integr Physiol 129, 65-73.

[2] Hildebran, J. N., Goerke, J. and Clements, J."Pulmonary surface film stability and composition" (1979) Journal of applied physiology 47, 604-611.

[3] Lang, C. J., Postle, A. D., Orgeig, S., Possmayer, F., Bernhard, W., Panda, A. K., Jürgens, K., Milsom, W. K., Nag, K. and Daniels, C. B. "Dipalmitoylphosphatidylcholine is not the major surfactant phospholipid species in all mammals" (2005) Am J Physiol Regul Integr Comp Physiol 289, R1426-39.

[4] Goerke, J., "Pulmonary surfactant: functions and molecular composition" (1998) Biochim Biophys Acta 1408, 79-89.

[5] McCormack, F. X. "Structure, processing and properties of surfactant protein A" (1998) Biochim Biophys Acta 1408, 109-131.

[6] Crouch, E. C. "Structure, biologic properties, and expression of surfactant protein D (SP-D)" (1998) Biochim Biophys Acta 1408, 278-289.

[7] Johansson, J. "Structure and properties of surfactant protein C" (1998) Biochim Biophys Acta 1408, 161-172.

[8] Hawgood, S., Derrick, M. and Poulain, F. "Structure and properties of surfactant protein B" (1998) Biochim Biophys Acta 1408, 150-160.

[9] Herting, E., Sun, B., Jarstrand, C., Curstedt, T., Robertson, B. "Surfactant improves lung function and mitigates bacterial growth in immature ventilated rabbits with experimentally induced neonatal group B streptococcal pneumonia" (1997) Arch Dis Child Fetal Neonatal Ed. 76(1), F3-8.

[10] Rauprich, P., Moller, O., Walter, G., Herting, E., Robertson, B. "Influence of modified natural or synthetic surfactant preparations on growth of bacteria causing infections in the neonatal period" (2000) Clin Diagn Lab Immunol. 7(5), 817-22.

[11] Oelberg, D. G., Xu, F. "Pulmonary surfactant proteins insert cation-permeable channels in planar bilayers" (2000) Mol Genet Metab. 70(4), 295-300.

[12] Munford, R. S., Sheppard, P. O. and O'Hara, P. J. "Saposin-like proteins (SAPLIP) carry out diverse functions on a common backbone structure" (1995) J Lipid Res 36, 1653-1663.

[13] Sarker, M., Waring, A. J., Walther, F. J., Keough, K. M. W. and Booth, V. "Structure of mini-B, a functional fragment of surfactant protein B, in detergent micelles" (2007) Biochemistry 46, 11047-11056.

[14] Sharifahmadian, M., Sarker, M., Palleboina, D., Waring, A. J., Walther, F. J., Morrow, M. R., Booth, V. "Role of the N-Terminal Seven Residues of Surfactant Protein B (SP-B)" (2013) PLoS ONE 8(9), e72821.

[15] Frey, S. L., Pocivavsek, L., Waring, A. J., Walther, F. J., Hernandez-Juviel, J. M., Ruchala, P., Lee, K. Y. C "Functional importance of the NH2-terminal insertion sequence of lung surfactant protein B" (2010) Am J Physiol Lung Cell Mol Physiol. 298(3), L335-L347.

[16] Walther, F. J., Waring, A. J., Hernandez-Juviel, J. M., Gordon, L. M., Wang, Z., Jung, C. L., Ruchala, P., Clark, A. P., Smith, W. M., Sharma, S., Notter, R. H. "Critical structural and functional roles for the N-terminal insertion sequence in surfactant protein B analogs" (2010) PLoS One 5, e8672.

[17] Johansson, J., Curstedt, T., Jörnvall, H. "Surfactant protein B: disulfide bridges, structural properties and kringle similarities" (1991) Biochemistry, 30 (28), pp 6917-6921.

[18] Olmeda, B., García-AÍvarez, B., Peŕez-Gil, J. "Structure-function correlations of pulmonary surfactant protein SP-B and the saposin-like family of proteins" (2013) Eur Biophys J 42, 209-222.

[19] Ahn, V. E., Leyko, P., Alattia, J. R., Chen, L. and Privé, G. G. "Crystal structures of saposins A and C" (2006) Protein Sci 15, 1849-1857.

[20] Ahn, V. E., Faull, K. F., Whitelegge, J. P., Fluharty, A. L. and Privé, G. G. "Crystal structure of saposin B reveals a dimeric shell for lipid binding" (2003) Proc Natl Acad Sci USA 100, 38-43.

[21] Rossmann, M., Schultz-Heienbrok, R., Behlke, J., Remmel, N., Alings, C., Sandhoff, K., Saenger, W., Maier, T. "Crystal structures of human saposins C and D: implications for lipid recognition and membrane interactions" (2008) Structure 16, 809-817.

[22] Hawkins, C. A., de Alba, E. and Tjandra, N. "Solution structure of human saposin C in a detergent environment" (2005) J Mol Biol 346, 1381-1392.

[23] de Alba, E., Weiler, S. and Tjandra, N. "Solution Structure of Human Saposin C: pH-Dependent Interaction with Phospholipid Vesicles" (2003) Biochemistry 42, 14729-14740.

[24] Liepinsh, E., Andersson, M., Ruysschaert, J. M. and Otting, G. "Saposin fold revealed by the NMR structure of NK-lysin" (1997) Nat Struct Biol 4, 793-795.

[25] Anderson, D. H., Sawaya, M. R., Cascio, D., Ernst, W., Modlin, R., Krensky, A. and Eisenberg, D. "Granulysin Crystal Structure and a Structure-derived Lytic Mechanism" (2003) J Mol Biol 325, 355-365.

[26] Leippe, M., Bruhn, H., Hecht, O. and Grotzinger, J. "Ancient weapons: the three-dimensional structure of amoebapore A" (2005) Trends Parasitology 21, 5-7.

[27] Hecht, O., Van Nuland, N. A., Schleinkofer, K., Dingley, A. J., Bruhn, H., Leippe, M. and Grotzinger, J. "Solution Structure of the Pore-forming Protein of Entamoeba histolytica" (2004) J Biol Chem 279, 17834-17841.

[28] Marrink, S. J., Risselada, H. J., Yefimov, S., , D. P., De Vries, A. H. "The MARTINI Force Field:? Coarse Grained Model for Biomolecular Simulations" (2007) J. Phys. Chem. B 111, 7812-7824.

[29] Marrink, S. J., De Vries, A. H., Mark, A. E. J. "Coarse Grained Model for Semiquantitative Lipid Simulations" (2004) Phys. Chem. B 108, 750-760.

[30] Monticelli, L., Kandasamy, S. K., Periole, X., Larson, R. G., Tieleman, D. P., Marrink, S. J. "The MARTINI Coarse-Grained Force Field: Extension to Proteins" (2008) J. Chem. Theory Comput. 4, 819-834.

[31] Altschul, S., Gish, W., Miller, W., Myers, E., Lipman, D. "Basic local alignment search tool" (1990) Journal of Molecular Biology 215 (3), 403-410.

[32] Hawgood, S., Derrick, M. and Poulain, M. "Structure and properties of surfactant protein B" (1998) Biochimica et Biophysica Acta 1408, 150-160.

[33] Andersson. M., Curstedt. T., Jornvall. H., Johansson. J. "An amphipathic helical motif common to tumourolytic polypeptide NK-lysin and pulmonary surfactant polypeptide SP-B" (1995) FEBS Lett 362, 328-332.

[34] Baoukina, S. and D. P., Tieleman "Direct Simulation of Protein-Mediated Vesicle Fusion: Lung Surfactant Protein B" (2010) Biophysical Journal 99(7) 2134-2142.

[35] Baoukina, S. and D. P., Tieleman "Lung Surfactant Protein SP-B Promotes Formation of Bilayer Reservoirs from Monolayer and Lipid Transfer between the Interface and Subphase" (2011) Biophysical Journal 100(7) 1678-1687.

[36] Scott, K. A., Bond, P. J., Ivetac, A., Chetwynd,A., P., Khalid,S., Sansom, M. S.P. "Coarse-grained MD simulations of membrane protein-bilayer self-assembly" (2008) Structure 16, 621-630.

[37] Ingólfsson, H. I., Melo, M. N., van Eerden, F. J., Arnarez, C., López, C. A., Wassenaar, T. A., Periole, X., De Vries, A. H., Tieleman, D. P. and Marrink, S. J. "Lipid Organization of the Plasma Membrane" (2014) J. Am. Chem. Soc. 2014, 136, 14554-14559.

[38] Rzepiela, A. J., Louhivuori, M., Peter, C., Marrink, S. J. "Hybrid simulations: combining atomistic and coarse-grained force fields using virtual sites" (2011) Phys. Chem. Chem. Phys. 13, 10437-10448.

[39] Shi, Q., Izvekov, S., Voth, G. A. "Mixed Atomistic and Coarse-Grained Molecular Dynamics: Simulation of a Membrane-Bound Ion Channel" (2006) J. Phys. Chem. B 110, 15045-15048.

[40] Han, W., Wan, C. K., Jiang, F., Wu, Y. D. "PACE Force Field for Protein Simulations. 1. Full Parameterization of Version 1 and Verification" (2010) J. Chem. Theory Comput. 6, 3373-3389.

[41] Han, W., Schulten, K. "Further Optimization of a Hybrid United-Atom and Coarse-Grained Force Field for Folding Simulations: Improved Backbone Hydration and Interactions between Charged Side Chains" (2012) J. Chem. Theory Comput. 8, 4413-4424.

[42] Zaltash, S., Palmblad, M., Curstedt, T., Johansson, J. and Persson, B. "Pulmonary surfactant protein B: a structural model and a functional analogue" (2000) Biochim Biophys Acta 1466, 179-186.

[43] Beck, D. C., Ikegami, M., Cheng-Lun, N., Zaltash, S., Johansson, J., Whitsett, J. A. and Weaver, T. E. "The Role of Homodimers in Surfactant Protein B Function in Vivo" (2000) J Biol Chem 275, 3365-3370.

[44] Zaltash, S., Griffiths, W., Beck, D. C., Duan, C. X., Weaver, T. E., Johansson J., "Membrane activity of (Cys48Ser) lung surfactant protein B increases with dimerisation" (2001) Biol. Chem., 382, 933-939.

[45] Olmeda, B., Garcia-Alvarez, B., Gomez, M. J., Martinez-Calle, M., Cruz, A. and Perez-Gil, J. "A model for the structure and mechanism of action of pulmonary surfactant protein B" (2015) The FASEB Journal 29,10, 4236-4247.

[46] Wan, C. K., Han, W., Wu, Y. D. "Parameterization of PACE Force Field for Membrane Environment and Simulation of Helical Peptides and Helix-Helix Association" (2012) J. Chem. Theory Comput. 8, 300-313.

[47] Andersson, M., Curstedt, T., Jornvall, H. and Johansson, J. "An amphipathic helical motif common to tumourolytic polypeptide NK-lysin and pulmonary surfactant polypeptide SP-B" (1995) FEBS Letters 362, 328-332.

[48] Palleboina, D., Waring, A. J., Notter, R. H., Booth, V., Morrow, M. "Effects of the lung surfactant protein B construct Mini-B on lipid bilayer order and topography" (2012) Eur Biophys J. 41, 755-767.

[49] Johansson, A., Curstedt, T. "Molecular Structures and Interactions of Pulmonary Surfactant Components" (1997) Eur. J. Biochem. 244, 675-693.

[50] Frishman, D., Argos, P. "Knowledge-Based Protein Secondary Structure Assignment" (1995) Proteins: Structure, Function, and Genetics 23, 566-579.

[51] Touw, G. W., Baakman, C., Black, J., AH te Beek, T., Krieger,E., Joosten, R. P., Vriend, G. "A series of PDB-related databanks for everyday needs" (2015) Nucleic Acids Research 43, D364-D368.

[52] Kabsch, W., Sander, C. "Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features" (1983) Biopolymers 22, 2577-2637.

[53] Guex, N. and Peitsch, M. C. "SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling" (1997) Electrophoresis 18, 2714-2723.

[54] Humphrey, W., Dalke, A. and Schulten, K. "VMD: visual molecular dynamics" (1996) J Mol Graph. 14, 33-38.

[55] VMD molecular visualization program introduction webpage:

http://www.ks.uiuc.edu/Research/vmd/ (accessed: 02/27/2014).

[56] General Purpose Cluster (GPC) of SciNet supercomputer centre:

http://www.scinethpc.ca/gpc/ (accessed: 09/12/2015).

[57] Khatami, M. H., Bromberek, M., Saika-Voivod, I., Booth, V. "Molecular dynamics simulations of histidine-containing cod antimicrobial peptide paralogs in self-assembled bilayers" (2014) Biochimica et Biophysica Acta 1838, 2778-2787.

[58] Qi, Y., Cheng, X., Han, W., Jo, S., Schulten, K. and Im, W. "CHARMM-GUI Micelle Builder for Pure/Mixed Micelle and Protein/Micelle Complex Systems" (2014) J. Chem. Inf. Model. 2014, 54, 1003-1009.

[59] Charm GUI webpage for lipid-bilayer system setup:

http://www.charmm-gui.org?doc=input/cgbilayer (accessed: 09/12/2015).

[60] Reva, B. A., Finkelstein, A., V., Skolnick, J. "What is the probability of a chance prediction of a protein structure with an rmsd of 6 Å?" (1998) Folding and Design 3 (2), 141-147.

[61] Maiorov, V. N., Crippen, G. M. "Significance of Root-Mean-Square Deviation in Comparing Three-dimentional Structures of Globular Proteins" (1994) J. Mol. Biol 235, 625-634.

[62] Web page for Wimley-White amino acid hydrophobicity scale:

http://blanco.biomol.uci.edu/hydrophobicity_scales.html.
(accessed: 09/12/2015)

[63] Bosshard, H. R., Marti, D. N., Jelesarov, I. "Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings" (2004) J. Mol. Recognit. 17, 1-16.

[64] Barlow, D. J., Thornton, J. M. "Ion-pairs in proteins" (1983) J. Mol. Biol. 168, 867-885.

[65] Introduction webpage for salt bridge plugin of VMD program:

http://www.ks.uiuc.edu/Research/vmd/plugins/saltbr/.

(accessed: 09/12/2015)

[66] Chernomordik, L. V., Kozlov, M. M. "Mechanics of membrane fusion" (2008) Nature Structural & Molecular Biology 15, 675-683.

[67] Stothard, P. "The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences." (2000) Biotechniques 28,1102-1104.

Chapter 4

Solid state NMR studies of pulmonary surfactant protein SP-B interacting with mechanically oriented lipid bilayers

4.1 Introduction

A mixture of lipids and proteins, i.e. lung surfactant, covers the air-water interface of the alveoli, which reduces the surface tension and helps us breathe. Of the proteins in lung surfactant, SP-B is the essential protein for breathing [1]. SP-B is from the saposin protein super-family and is thus predicted to have 4 to 5 helical regions [2, 3]. SP-B has the highest hydrophobicity among saposin super family proteins **(Table 3.1)**. The exceptionally high hydrophobicity of SP-B makes it difficult

to study experimentally, even with techniques that have been shown to work with other hydrophobic proteins such as membrane proteins. Thus far, there is no atomic resolution experimental structure for SP-B. However, in Chapter 3, I present all-atom molecular dynamics (MD) simulations of SP-B interacting with a lipid bilayer. In this chapter, I present steps towards combining the MD simulations with experimental data to provide a high quality structural model that is consistent with the limited experimental data it is possible to collect on SP-B.

X-ray crystallography and NMR spectroscopy are the two main techniques to elucidate the structure of proteins. X-ray crystallography has been used to determine the structure of some of the less hydrophobic saposin super-family members [5–8]. With this method it is necessary to grow crystals of the protein, which is difficult, if not impossible, for extremely hydrophobic proteins like SP-B. Besides X-ray crystallography, people also use solution NMR to determine the structure of membrane proteins, including some of saposin super-family members, in detergent micelles [10–16, 19]. Even though fragments of SP-B, e.g. Mini-B, have been studied employing solution NMR [17, 19, 22], nevertheless, the structure of full length SP-B has so far proved intractable for solution NMR. No one has been able to get any reasonable solution NMR spectra due to either aggregation and/or conformational inhomogeneity.

Circular dichroism and infrared spectroscopy, while unable to provide information on the overall 3D structure, do indicate overall secondary structure, and have been used to show that SP-B has an overall 35-50% helicity [20–22]. Moreover, the homology modelling of SP-B, based on saposin super-family proteins, predict 4-5 helical regions for the protein [25, 26]. Sarker et al. [17] showed that Mini-B, a construct of SP-B residues 8-25 and 63-78, has two helical regions. Recently, Olmeda et al. [27]

fit a homology model of SP-B into a low resolution structural envelope of SP-B, which was derived from the atomic force microscopy, in combination with electron microscopy. They proposed a tube-like multimer model for the SP-B structure.

Unlike solution NMR, which is limited to small protein/detergent complexes, solid state NMR is a powerful method to study the most biologically relevant systems for membrane proteins, e.g. lipid bilayers and proteins. Researchers have used ^{2}H, ^{31}P and ^{15}N solid state NMR to understand the position and the orientation of the SP-B fragments in lipid bilayers [22–24]. Several studies have shown that ^{15}N solid state NMR is a practical method to study the conformation and orientation of proteins with relatively large helical regions in oriented lipid bilayers. For example, Kandasamy et al. [29] studied the role of oligomerization in the structure of the second transmembrane helix of $GABA_A$ protein, using selectively ^{15}N-labeled protein. Kawaguchi et al. [29] studied the structure and orientation of the two amphipathic α-helical regions in CAMA antimicrobial peptide, interacting with lipid bilayers. There are other studies, in which people have studied larger proteins employing solid-sate NMR. For example, Aisenbrey et al. [30] employed 1D ^{15}N solid state NMR to determine the orientation of two α-helical regions in 209 residue Bcl-X_L protein, using selectively ^{15}N-labeled protein. Moreover, Vosegaard et al. [31] used selective labeling of ^{15}N-methionine, ^{15}N-valine and ^{15}N-glycine in a 7-helix transmembrane protein to find the tilt angle of each helix.

In general, these oriented sample ^{15}N solid state NMR methods are applied by first calculating the theoretical NMR spectra of the protein's helical regions for different orientations within the bilayer. Then, the simulated spectra are compared with the actual experimental results to indicate the actual structural parameters. My goal is to

use the SP-B structures derived from my all-atom MD simulations in Chap 3 to help to determine the actual experimental structure of SP-B. To do this, I will predict the ^{15}N solid state NMR spectra from the computational structures employing theoretical methods. Then, I need to compare the theoretical spectra with the experimental spectra of ^{15}N-labeled protein in oriented lipid bilayers. The best fit between the computed spectra and the actual spectrum should indicate the closest computational structure to the NMR result.

The Hamiltonian for the chemical shift interaction in solid state NMR can be written as $H_{CS}=\gamma \mathbf{\bar{I}}.\boldsymbol{\sigma}.\mathbf{\bar{B}_0}$, where γ is the gyromagnetic ratio, $\mathbf{I}$ is the spin vector, $\boldsymbol{\sigma}$ is the chemical shift tensor and $\mathbf{B_0}$ is the external magnetic field. As presented by the Bechinger group [28, 34–36], in general, one can write the anisotropic chemical shift interaction of a ^{15}N nucleus as a second-rank tensor:

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{pmatrix}.$$

We can find a coordinate system, named the principal axis system (PAS), where only the diagonal elements of the transformed tensor (σ_{11}, σ_{22} and σ_{33}) remain **(Figure 4.1-A)**. The vector σ_{33} points along the z' axis of the PAS, which in turn is parallel to the long axis of an ideal helix. The zz element of the chemical shift tensor in the laboratory frame corresponds to the measured value of chemical shift in an NMR experiment, assuming that B_0 points in the z direction. Employing Euler angles Φ, Θ and Ψ we can transform coordinates from the lab frame to the PAS with

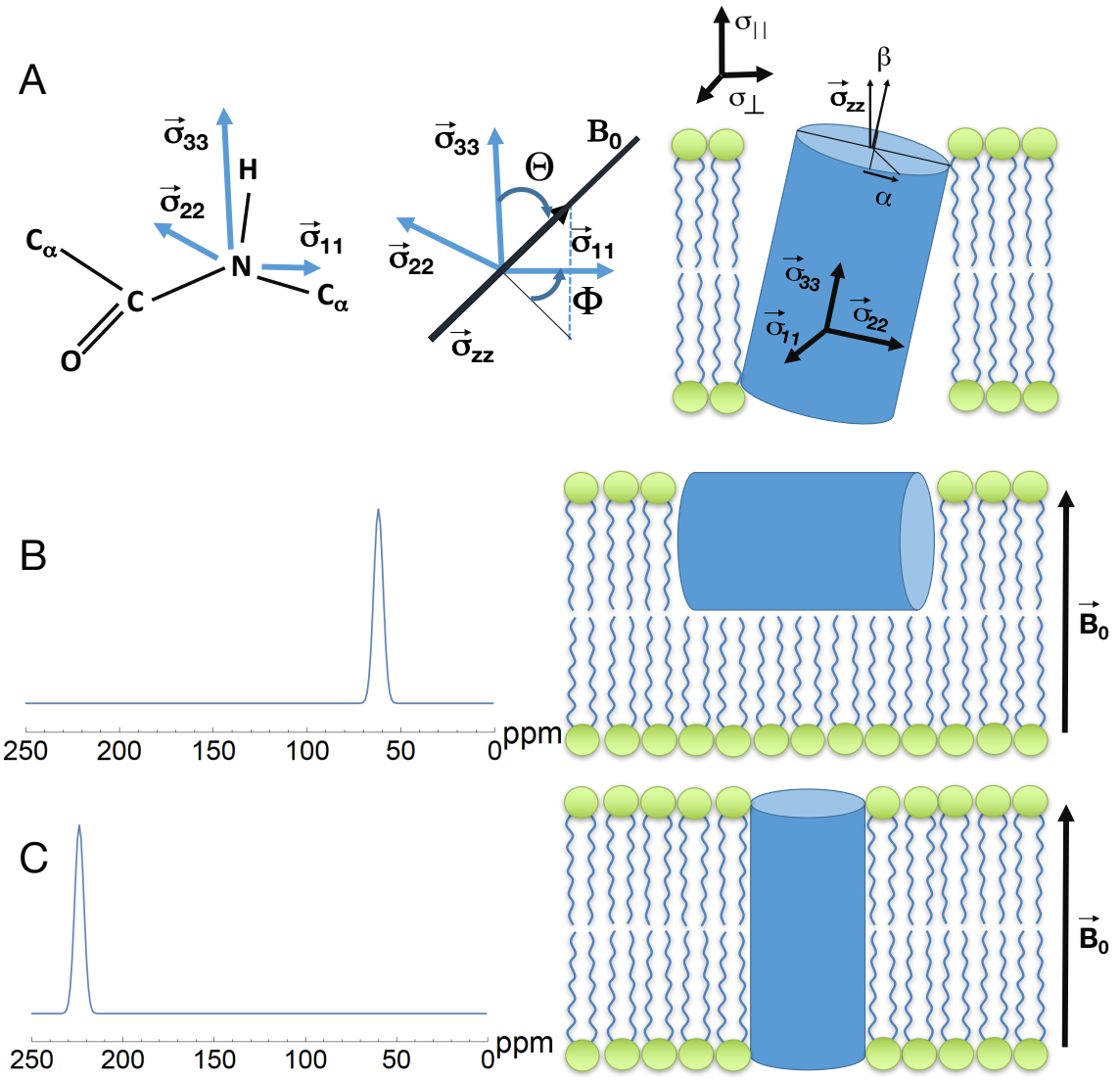

Figure 4.1: A) The representation of the ^{15}N chemical shift tensor for a ^{15}N nucleus in the peptide bond and the angles Θ and Φ, as well as α and β for a helical region of the peptide (cylinder) in an oriented lipid bilayer. Θ and Φ match α and β when the bilayer normal and the long axis of the cylinder. On the left, σ_{11} and σ_{33} are in the plane of the paper, while σ_{22} is into the paper. Panels B and C show the ^{15}N spectra expected for a helix with a helical axis perpendicular to the bilayer normal, and for a helix with helical axis parallel to the bilayer normal. Figure is inspired from Ref. 28

the rotation matrices:

$$\mathbf{R}_z(\Phi) = \begin{pmatrix} \cos\Phi & \sin\Phi & 0 \\ -\sin\Phi & \cos\Phi & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{R}_{y'}(\Theta) = \begin{pmatrix} \cos\Theta & 0 & -\sin\Theta \\ 0 & 1 & 0 \\ \sin\Theta & 0 & \cos\Theta \end{pmatrix},$$

$$\mathbf{R}_{z''}(\Psi) = \begin{pmatrix} \cos\Psi & \sin\Psi & 0 \\ -\sin\Psi & \cos\Psi & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

giving $\boldsymbol{\sigma'}_{lab} = \boldsymbol{R}\,\boldsymbol{\sigma}_{PAS}\boldsymbol{R^{-1}}$, where $\boldsymbol{R} = \mathbf{R}_{z''}(\Psi)\mathbf{R}_{y'}(\Theta)\mathbf{R}_z(\Phi)$.

Thus, we obtain σ_{zz} in terms of the PAS chemical shift components:

$$\sigma_{zz} = \sigma_{11}\sin^2\Theta\cos^2\Phi + \sigma_{22}\sin^2\Theta\sin^2\Phi + \sigma_{33}\cos^2\Theta. \tag{4.1}$$

The chemical shift values of σ_{11} and σ_{22} are approximately 65 ppm and 85 ppm respectively, while the chemical shift value of σ_{33} is quite different and is in the vicinity of 230 ppm. With the values of σ_{11} and σ_{22} being similar (compared to σ_{33}), the tensor nearly possesses cylindrical symmetry. In the approximation that σ_{11} and σ_{22} are equal, we can rewrite **Equation 4.1** as [28],

$$\sigma_{zz} = (\bar{\sigma}_{\parallel} - \bar{\sigma}_{\perp})\cos^2\Theta + \bar{\sigma}_{\perp}, \tag{4.2}$$

where, $\bar{\sigma}_{\perp} = (\sigma_{11} + \sigma_{22})/2$ and $\bar{\sigma}_{\parallel} = \sigma_{33}$ and Θ corresponds to the angle between the helix long axis, which is assumed to be parallel to σ_{33}, and the bilayer normal.

This cylindrical symmetry can also be obtained e.g., if one can average over Φ in the case that there is no preferred value of Φ. Equation 4.2 shows that if the peak in the ^{15}N spectrum is located above 200 ppm, this means that the α-helical region of the protein is in a more transmembrane position with respect to the lipid bilayer, i.e. Θ roughly equals to zero **(Figure 4.1-B)**. Whereas, if the ^{15}N spectrum of the protein has a peak located below 100 ppm, the α-helical region is expected to be in a parallel orientation with respect to the lipid bilayer **(Figure 4.1-C)**.

In a more realistic calculation, we should consider the liquid crystalline nature of the lipid membrane and the dynamics of the protein. Here, we need to take into account the rotational averaging of the protein around the bilayer normal that occurs in the membrane due to the continuous diffusion of the protein within the lipid membrane [28]. This averaging generates an axially symmetric tensor, in which $\sigma_\perp$ and $\sigma_\parallel$ are the perpendicular and parallel components of the chemical shift tensor (perpendicular and parallel with respect to the bilayer normal). The values of $\sigma_\parallel$ and $\sigma_\perp$ will be different from $\bar{\sigma}_\parallel$ and $\bar{\sigma}_\perp$ because we are averaging around the bilayer normal not the axis of the helix. We now introduce the Euler angles of α (pitch) and β (tilt), explained in **Figure 4.1-A**, with respect to the bilayer normal. We can express $\sigma_\parallel$ and $\sigma_\perp$ in terms of α and β as:

$$\sigma_\parallel = \sigma_{11}\cos^2\alpha\sin^2\beta + \sigma_{22}\sin^2\alpha\sin^2\beta + \sigma_{33}\cos^2\beta,$$

$$\sigma_\perp = \frac{1}{2}[\sigma_{11}(1-\cos^2\alpha\sin^2\beta) + \sigma_{22}(1-\sin^2\alpha\sin^2\beta) + \sigma_{33}\sin^2\beta].$$

With $\sigma_{xx} = (\sigma_{11}-\sigma_{22})\cos^2\alpha + \sigma_{22}$, we have:

$$\sigma_\parallel = (\sigma_{33}-\sigma_{xx})\cos^2\beta + \sigma_{xx}, \tag{4.3}$$

$$\sigma_{\perp} = \frac{1}{2}[(\sigma_{11} + \sigma_{22}) + (\sigma_{33} - \sigma_{xx}) \sin^2 \beta], \tag{4.4}$$

where we can see the resemblance of **Equation 4.3** with **Equation 4.2**. In the case of the bilayer normal being parallel to the external magnetic field, $\sigma_{\parallel}$ matches σ_{zz} and $\sigma_{\parallel}$ is the experimentally measured value. For β=0°, $\sigma_{\parallel}$=σ_{33} and for β=90°, $\sigma_{\parallel}$=$\sigma_{xx} = (\sigma_{11} - \sigma_{22}) \cos^2 \alpha + \sigma_{22}$, which is a value between σ_{11} and σ_{22}. These expressions for $\sigma_{\parallel}$ and $\sigma_{\perp}$ now can be used instead of $\bar{\sigma}_{\parallel}$ and $\bar{\sigma}_{\perp}$ in **Equation 4.2**, where Θ now represents the angle between the bilayer normal and B_0.

In order to apply these expressions to the chemical shift of a ^{15}N nucleus in the helical portion of the protein, one must acquire the ^{15}N NMR spectrum of that region. Then, the standard way to analyze the ^{15}N spectrum is to start with a PDB structure file of the helical region - either a homology model or an ideal helix. Next, the researchers find the orientation of the specified residue with respect to the helical region of the protein. In this method, they rotate the protein's structure, i.e. the PDB file, around two axes, the pitch axis, the central axis of the helix structure (pitch angle) and the tilt axis, the axis perpendicular to pitch axis (tilt angle) [30]. This will screen all the possible 180×180 degree angles of tilt and pitch angles of helix structure for the acquired chemical shift value.

My plan was to implement these strategies a little differently; I had planned to calculate the expected ^{15}N spectrum, not for all the possible orientations of tilt and pitch - but using the structures from my MD simulations. An important point in these calculations is that, since unstructured residues, as well as interhelical loops, possess no structural information in the ^{15}N solid state NMR studies of oriented samples [43], thus, there is no data pertaining to these parts in the predicted spectra.

In this chapter, the goal is to combine the solid state NMR of uniformly ^{15}N-labeled SP-B in a mechanically oriented lipid bilayer with the computational simulations from Chapter 3 in order to gain information about the orientation of the helical regions of the protein with respect to the lipid membrane. Firstly, I write the computational code to simulate the ^{15}N spectra for each MD simulation. Next, I prepare oriented bilayers with and without SP-B and use ^{2}H and ^{31}P NMR to check the sample orientation. Finally, I prepare oriented bilayers with uniformly ^{15}N-labelled SP-B and run ^{15}N NMR experiments.

4.2 Materials and Methods

4.2.1 Prediction of ^{15}N Chemical Shifts in NMR Spectra of Oriented Samples

In this work, I follow the approach used by the Bechinger group to calculate the expected ^{15}N spectrum of the α-helical regions of SP-B from my MD simulations. In this method, the orientation of each α-helical region in the protein provides a defined ^{15}N NMR peak in the spectra [28].

In Chapter 3, I presented the all-atom MD simulations of SP-B interacting with lipid bilayer systems. Since I use uniformly ^{15}N-labeled protein and acquire 1D NMR signal with bilayer normal parallel to the B_0, we can not specifically distinguish the source residue of the signal experimentally. Thus, in an ideal case, we can calculate the tilt angle of the helical region with respect to the lipid bilayer and we use **Equation 4.2**. The essential part of calculating the spectra in this method is to find the

orientation of the NH bond in the protein's backbone. First, I calculate and store positions of N and H atoms of the residues in the α-helical regions of the structure. Then I use the N and H position values to determine the NH vector. I use this vector as an input file for the code, written in MATHEMATICA, to predict the ^{15}N spectra using **Equation 4.2**. First, I normalize the NH vector and calculate the Θ and Φ (or equivalently α and β, while the bilayer normal is parallel to B_0) angles of the NH vector. I calculate these angles for each frame for the last 200 ns of the simulation (200 frames in total). The chemical shift values for the NH bond are 75 ppm, 61 ppm and 223 ppm for σ_{11}, σ_{22} and σ_{33} respectively [39]. To introduce the line broadening that will be present in the experimental solid state NMR spectra, I calculate the observed chemical shifts for a Gaussian distribution of σ_{zz} around the predicted σ_{zz} values, to produce experiment-like ^{15}N spectra. I chose the standard deviation of 2.5 in the distribution function, however one can try different values to match the plot with the experimental results. In my code, the external magnetic field is set in the Z direction parallel to the bilayer normal. Details of the code, implementing **Equation 4.2**, are in Appendix B.

4.2.2 SP-B Preparation[1]

We employ recombinant expression to produce SP-B. In this method, we introduce DNA coding for SP-B into bacteria, which produce the protein. To uniformly label SP-B with ^{15}N, we grow the bacteria with only one nitrogen source $^{15}NH_4Cl$.

Our group has optimized the SP-B preparation protocol for refolding SP-B into

[1]SP-B preparation and purification were carried out primarily by Ms. Donna Jackman, Booth lab.

detergent micelles. However, for my solid state NMR experiments, the protein was refolded into lipids, i.e. POPC/POPG with 9:1 ratio. In general, we confirm the presence of SP-B by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). However, I also performed ^{15}N solid state NMR with the samples at various stages of preparation in order to determine how much ^{15}N signal I was getting from the samples. In the last step of purification, we "desalted" the SP-B in order to remove any salt, detergent, or other non-tightly bound molecules from the sample.

4.2.3 Mechanically Oriented Lipid Bilayers

To produce mechanically oriented bilayers, I use a method similar to the one described by Rainey et al. [32], in which they use mica slides to support lipid bilayers and keep their alignment during the experiment. In **Figure 4.2** we can see the different steps of sample preparation. I purchase the Mica Grade-I in packs of 20 sheets from SPI Supplies®. I cleave several layers from both sides of the mica sheet, using a normal scalpel, which gives fresh mica surfaces on both sides. Then, I cut the sheet in the size needed to fit inside the coil of the NMR probe, e.g. 12.5 mm×5 mm in my case **(Figure 4.2-A)**. To cut the plates, I use a scalpel. While cutting the mica plates, one should avoid cracks and defects, which appear on mica sheet after cleaving it. One should note that cutting plates from thick sheets produce curved and flaky edges. I always cut an extra plate to be used as a cap to my plate stack. Finally, I clean the mica filings off the plate's surface, using air flow, then I put them in a petri dish until depositing my sample on them.

After preparing the mica plates, I prepare my stock solution to deposit on each

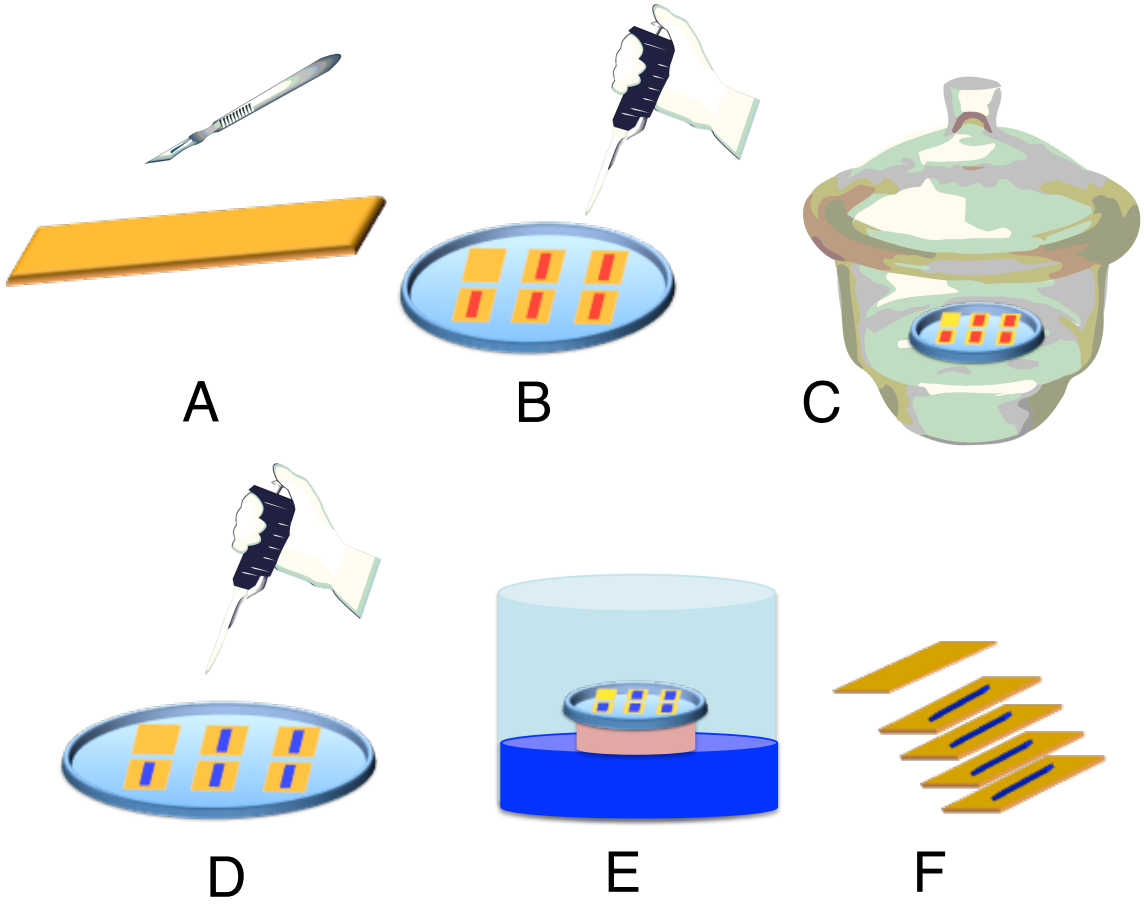

Figure 4.2: Steps of sample preparation. A) Cleaving and cutting mica sheets into mica plates, B) Depositing stock solution on mica plates, C) Drying out the samples, D) Depositing deuterium-depleted water on my dried samples, E) Putting samples in hydration chamber and F) stacking the mica plates and wrap them to prepare my NMR sample.

plate. Depending on the type of the experiment, I use slightly different methods to prepare my stock solutions. For my thesis, I run two experiments, one to prepare a mechanically oriented lipid bilayers sample using pure POPC and the other to prepare mechanically oriented lipid bilayers containing SP-B sample, i.e. SP-B plus the POPC/POPG lipid associated with it as part of the refolding and purification process, added to deuterated POPC. To prepare the stock solution for the pure POPC

sample, I dissolve 5 mg of POPC lipid (POPC:$POPC_{d31}$) with 3:7 mass ratio, in 100 μl of chloroform:methanol (C:M) with 1:1 volume ratio. For the next one, the SP-B+POPC/POPG+$POPC_{d31}$ stock solution, I use more steps as the SP-B does not dissolve in C:M with 1:1 volume ratio. Normally, for small peptides in oriented bilayers, we prepare a stock solution of protein+lipid by co-dissolving them in C:M solvent. However, this protocol had to be modified for SP-B which is not soluble in C:M. Thus, I dissolve 2.5 mg of the protein sample, i.e. SP-B+POPC/POPG, in 3 ml methanol. The standard procedure of preparing SP-B in detergents produces 0.4–2 mg ($\sim$1.2 mg on average) of SP-B in approximately 12 mg of SP-B+detergents, i.e. 10% of the sample is pure SP-B. Thus, based on my calculations, if I have the same ratio of protein to lipid in my SP-B sample, 2.5 mg of the SP-B sample contains $\sim$0.25 mg of SP-B and $\sim$2.25 mg of lipids, which are tightly bounded to the SP-B. Separately, I dissolve 12 mg of $POPC_{d31}$ in 100 μl of C:M with 1:1 volume ratio (similar to the pure POPC sample). Finally, I mix the 2.5 mg of SP-B+POPC/POPG stock solution with the 12 mg of $POPC_{d31}$ stock solution, which gives me $\sim$0.2 mol% ratio of protein to lipid. For both the pure POPC and the SP-B+POPC/POPG+$POPC_{d31}$ systems, the stock solutions should be clear with no observable solid particles in them. These stock solutions are less concentrated than is normally used for making mechanically oriented bilayers and are hard to deposit on the plates. Therefore I concentrate the stock solutions under a stream of nitrogen gas, until they are easy to work with.

In the next step, I deposit the stock solution on the mica plates, using a micro pipette with a 10 μl pipette tip **(Figure 4.2-B)**. On each round of depositing the stock solution, I add about 0.5 μl of the solution, as a thick line in the centre of the rectangular plates. I need to prevent it from going near the edges, which could ruin

my sample. After each round, I wait for a minute or two for the plates to partially dry, which helps to control the newly poured material in the middle of each plate. At the end of this step, which may take up to 3 hours, I have my mica plates covered with a thick-line-shaped pile of partially dried materials. One should prevent air bubbles being trapped inside the dried material and keep the cap plate blank. In each experiment, each plate should have 0.5-1.2 mg of the sample. Exceeding the 1.2 mg of the sample can increase the risk of the sample over flow on the edges and lead to randomly oriented lipid bilayers. On the other hand, using less than 0.5 mg of the material on each plate will prevent the mica plates from fully contacting and thus, the lipid between them can not make the mechanically oriented structure.

Then, we need to dry out the samples, completely. Here, I put my samples under the fume hood for about an hour. Then, I put the plates into the vacuum chamber for overnight (about 15-20 hours) to completely evaporate the solvent out of the sample **(Figure 4.2-C)**.

Now, I hydrate the sample with deuterium-depleted water. I add 0.5 μl of water to each dried pile of material on the mica plates **(Figure 4.2-D)**. I need to wait approximately 3 minutes after each round of water deposition to let my sample absorb the water. Similar to depositing the stock solution on mica plates, one should prevent the material running towards the edges of plates. I continue these steps until each plate gets 3-4 μl of deuterium-depleted water. I then added each mica plate to a hydration chamber. To prepare the chamber, I pour $\sim$200 ml of distilled water in the bath and add 65 gr of ammonium phosphate dibasic, as well as a small amount of sodium azide to prevent bacterial growth **(Figure 4.2-E)**. The amount of water to add to the mica plates, as well as the duration of keeping them inside the chamber

depends on the sample contents. However, to get the full hydration, the samples should be in the chamber for a minimum of 24 hours.

In the final step of sample preparation, I need to stack and wrap the mica plates, which must be performed with great care in order to achieve good orientation **(Figure 4.2-F)**. In this step, after taking the samples out of the hydration chamber, I start placing plates on each other, using a fine tip tweezer. One should prevent the plates sliding over each other. I make stacks of 4-5 mica plates with an additional plate used as a cap. Then, I use plastic wrap to seal each stack.

After covering my samples in 3-4 rounds of plastic wrap, I put them in my NMR tube (one or more stacks depending on the experiment). Then, I add about 1 μl of deuterium-depleted water to prevent the sample from drying out (but note that too much water will destroy the sample). Then I seal the NMR tube with paraffin wax and keep in the fridge for at least 24 hours before doing the experiment, which allows the lipid bilayers to become further oriented. I divide my SP-B+POPC/POPG+POPC$_{d31}$ sample into 3 packs, each containing 5 stacks of mica plates. Then, I wrap them individually. After running ^{2}H NMR, I keep the mechanically oriented samples to do ^{15}N NMR experiment.

4.2.4 NMR Acquisition

All of the experiments are performed at 298 K. In the mechanically oriented samples, the bilayer normal is perpendicular to the magnetic field. The ^{2}H experiments are performed on Dr. Morrow's locally-assembled 9.4 T spectrometer, using a cylindrical coil. In deuterium experiments, we applied a quadrupolar echo sequence of ($\pi/2_y$ –

$\tau - \pi/2_x - \tau$) to avoid problems with preamplifier recovery time [33]. The τ is set to 30 μs with a pulse length of 4-4.25 μs with a recycle delay of 0.9 s.

Both ^{31}P and ^{15}N experiments are performed on a Bruker Avance II 14.1 T (600 MHz) spectrometer in the Centre for Chemical Analysis, Research and Training (C-CART) facility, employing a Bruker 4 mm triple double-tuned flat coil probe HN(P). The 1D ^{31}P solid state NMR spectrum in **Figure 4.5** is acquired at 243.01 MHz. 85% H_3PO_4 is used to reference the phosphorous chemical shift frequencies at 0 ppm, externally. The spin echo pulse sequence in our experiments is ($\pi/2 - \tau - \pi-$acq) with the $\pi/2$ pulse length of 12 μs, τ of 22 μs and a recycle delay of 3 s with a 62.5kHz high-power decoupling pulse, used during the acquisition. For the ^{15}N experiments, we used cross polarization (CP) experiments, performed in a 62.5 kHz B_1 field. The contact and the recycle delay times are 1.5 ms and 5 s respectively. Here, we acquire the signal using high power ^{1}H decoupling pulses. In **Figure 4.6-E** the bilayer normal is parallel to the external magnetic field. We set glycine as the reference to 35.7 ppm, which corresponds to 0 ppm of the liquid NH_3.

4.3 Results

4.3.1 Predicted Spectra

In chapter 3, we discussed the results we got from the computational simulations of SP-B interacting with a POPC bilayer. Having the coordinates of the N and associated H atoms in all-atom simulations, we can predict the ^{15}N spectra of SP-B. In **Figure 4.3**, we can see the predicted NMR spectra for SP-B, as a summation of

every calculated spectrum over the last 200 ns of each simulation (20 configurations). I examined the behaviour of the predicted NMR spectra over the last 200 ns of the simulation using a script (copied in Appendix B) that displays an animation of the predicted ^{15}N spectrum for each snapshot. During the simulation, tiny changes in the secondary structure of the protein in each snapshot slightly alter the intensity of the spectra. Besides, tiny changes in the orientation of the α-helical domains, i.e. changes in the tilt angle of the helix with respect to the bilayer normal, vary the position of the peaks in the predicted spectra of the protein. In my simulations, as the overall orientation of helical regions (not the whole protein) are mainly parallel to the lipid bilayer, the predicted spectra look similar to each other. As expected for helices that are generally positioned in parallel with the lipid surface, the spectra have most of their intensity below 100 ppm.

4.3.2 Experimental Results

The standard protocol used in the lab to prepare SP-B experimentally produces SP-B in detergent micelles. However, detergent micelles are not compatible with the solid state NMR method I need to employ, which requires that the protein be in mechanically oriented lipid bilayers. Therefore, I needed to use SP-B produced via a modified protocol that resulted in SP-B being refolded and purified in POPC/POPG with 9:1 ratio (POPC/POPG is associated with SP-B). A second modification to the standard SP-B production method was also needed. Normally, oriented solid state NMR samples are produced via co-dissolving the dry protein and dry lipid components in an organic solvent such as a mixture of chloroform and methanol. Thus, my SP-B

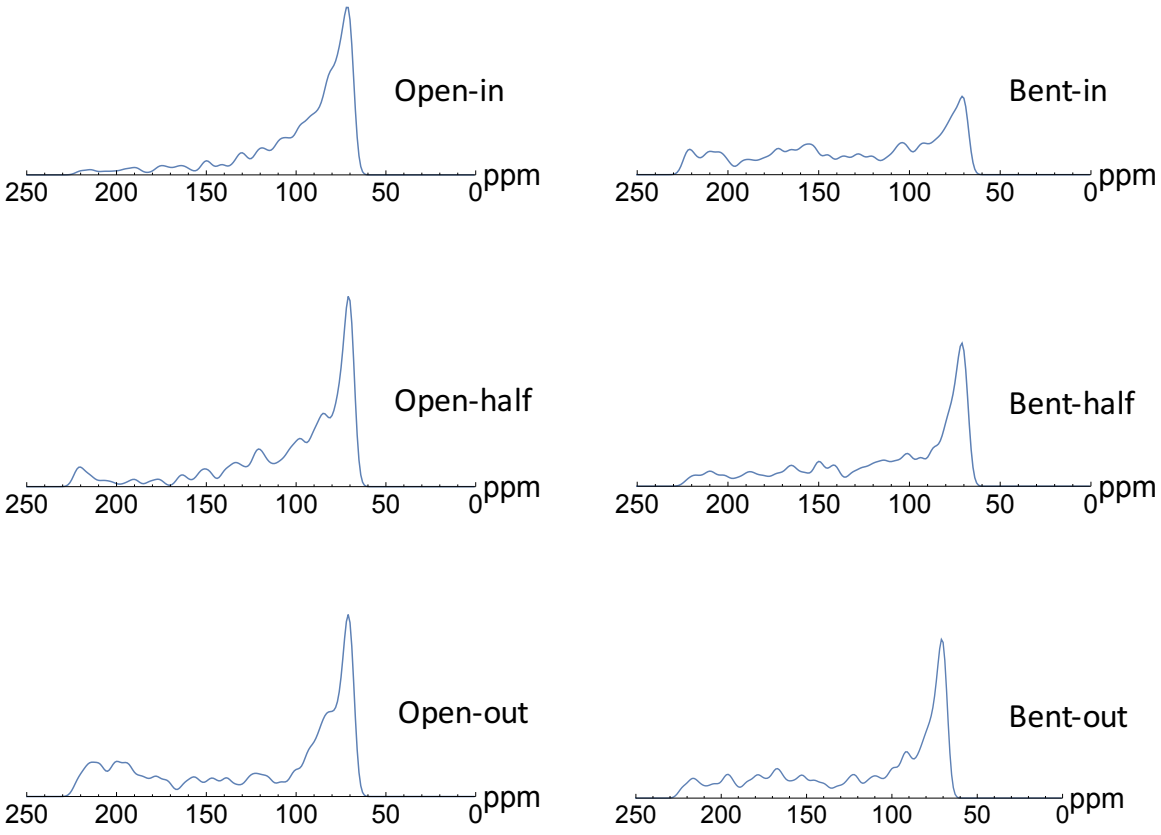

Figure 4.3: Predicted ^{15}N solid state NMR spectra of all-atom MD simulations of SP-B configurations in oriented POPC lipid bilayer, presented in Chapter 3. Each spectrum is calculated for 20 configurations sampled over the last 200 ns of simulation (1 configuration per 10 ns). The scaling in the vertical direction is in the same for each panel.

protein (in POPC/POPG) needed to be dried via lyophilization (freeze-drying). The protein and lipid contents of my SP-B sample were estimated at ∼0.25 mg of SP-B and ∼2.25 mg of POPC/POPG (9:1).

Before collecting ^{15}N data on oriented SP-B samples, I first need to practice with the pure POPC to enhance my experience in orienting the lipid bilayers between mica plates. This practice was important to ensure that any loss of orientation observed for the samples with SP-B was due to the protein and not to an inadequacy in the sample preparation. I use ^{2}H solid state NMR experiments, employing a locally-assembled 9.4 T spectrometer, to study the orientation of the lipid bilayers in my POPC lipid bilayer samples. **Figure 4.4-A** represents the deuterium spectrum of mechanically oriented lipid bilayers of a pure POPC sample for an hour of scanning, i.e. 4000 scans. This sample contains 5 mg of POPC/POPC$_{d31}$ (3:7), deposited on 5 rectangle mica plates.

As a second step I needed to ensure that I could obtain well oriented samples in the presence of SP-B, as well. **Figure 4.4-B** represents the ^{2}H NMR spectrum of the mechanically oriented bilayer containing SP-B. Here, the lipid bilayers show less orientation compared to the SP-B-free sample in **Figure 4.4-A**, indicating the effect of SP-B on the bilayers. However, based on the position of peaks in the ^{2}H spectrum (**Figure 4.4-B**) and the results of the ^{31}P spectrum in **Figure 4.5** (discussed in the following text), the amount of perturbed lipid in the system is not so much, as to ruin the sample for ^{15}N NMR experiments. I made 3 similar stacks of SP-B+POPC/POPG+POPC$_{d31}$ sample. However, only one of the three samples formed well oriented bilayers, while the other two were randomly oriented due to mishandling the sample in the wrapping step. The sample shown in **Figure 4.4-B** contains the

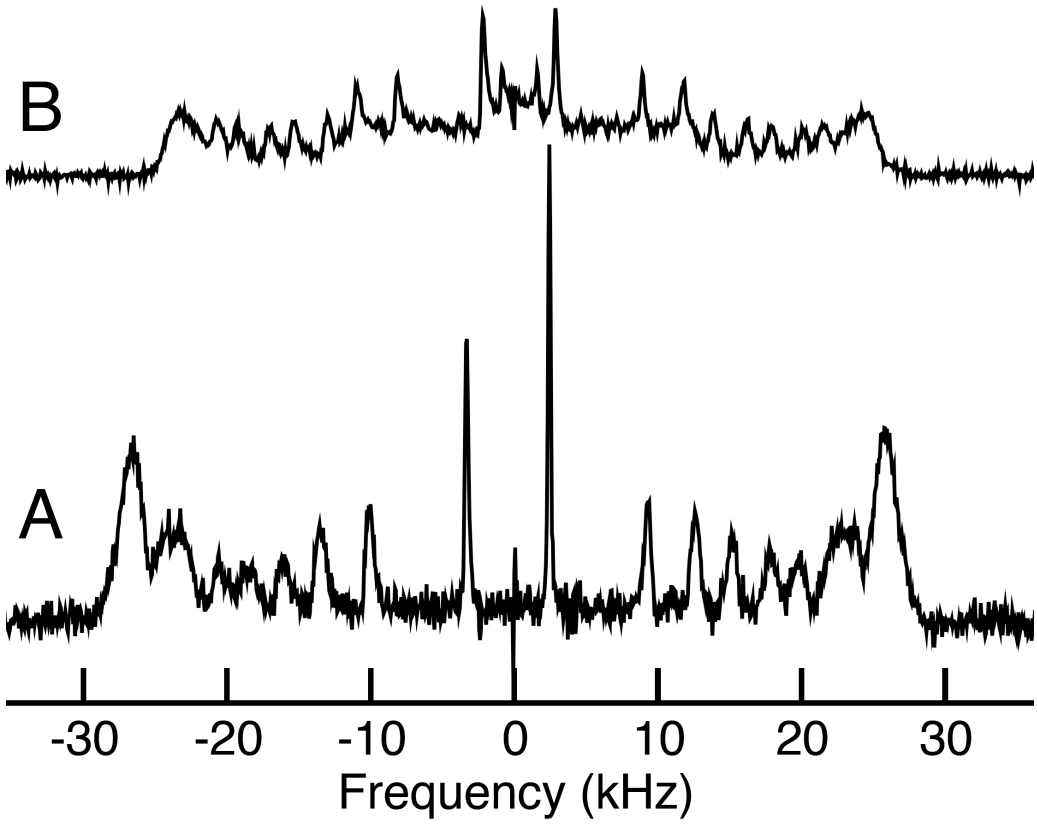

Figure 4.4: The deuterium spectra of lipid bilayers. A) a mechanically oriented POPC lipid bilayer, B) mechanically oriented POPC lipid bilayer containing ∼0.8 mg of SP-B sample and 4 mg of $POPC_{d31}$, i.e. the SP-B+POPC/POPG+$POPC_{d31}$ sample. The spectra were acquired with 4000 transients for A, which contains 5 mg of $POPC_{d31}$ and 16000 transients for B, which contains 4 mg of $POPC_{d31}$ at 298 K, employing a locally-assembled 9.4 T spectrometer. Vertical scale is the same for both spectra.

^{15}N-labelled SP-B plus the unlabeled lipids that bound to SP-B during the refolding process, as well as the deuterated lipids I added to the sample. The 0.8 mg of SP-B+POPC/POPG is estimated to contain ∼0.1 mg of SP-B and ∼ 0.7 mg of POPG/POPG (9:1) and to this I added 4 mg of POPC$_{d31}$ (12 mg for 3 samples, thus, 4 mg for one sample), deposited over 4 rectangle mica plates. The spectrum is acquired with 16000 transients. The spectrum of **Figure 4.4-A** is typical of a well oriented lipid sample, while in the presence of SP-B, in **Figure 4.4-B**, the spectrum is consistent with a superposition of spectra: a well oriented portion, similar to panel A, superimposed with a spectrum typical of a randomly oriented sample at half the oriented splittings. I have tried this experiment with higher concentrations of protein/POPC$_{d31}$, but these samples were not well oriented. This could be due to the fact that SP-B perturbs lipid bilayers at the higher concentrations of protein/lipid.

In addition to ^{2}H spectra, I use the ^{31}P and ^{15}N spectra, employing a Bruker Avance II 14.1 T spectrometer to study my SP-B+POPC/POPG+POPC$_{d31}$ sample. To ensure the sample is oriented after my ^{15}N NMR experiment, I run the ^{31}P spectra of the SP-B+POPC/POPG+POPC$_{d31}$ system **(Figure 4.5)**. The peak at 37.4 ppm represents oriented lipid bilayers perpendicular to the external magnetic field. The small randomly oriented part in the spectrum (∼-12 ppm) could be due to the influence of the protein on lipid bilayers by slightly perturbing the mechanical orientation of the system. The spectrum we acquired from the highly oriented sample indicates that this sample retains its orientation at room temperature after 6 days of running an ^{15}N experiment.

In my initial ^{15}N NMR experiment I acquired 110000 scans with a sample estimated to contain almost 0.1 mg of ^{15}N labelled SP-B (0.8 mg of SP-B+ POPC/POPG

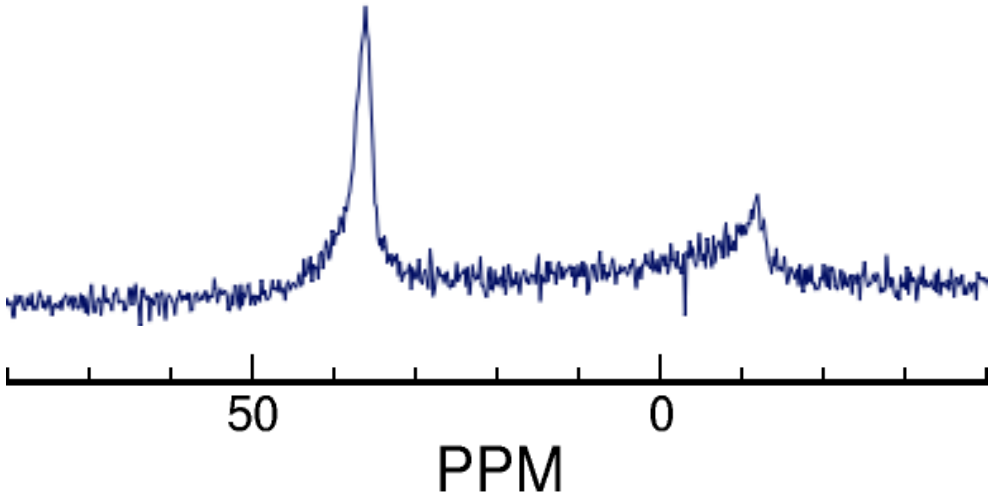

Figure 4.5: The ^{31}P spectrum of the POPC lipid bilayer in SP-B+POPC/POPG+POPC$_{d31}$ system. We can see the main portion of the lipid bilayer is oriented (peak value of 37.4 ppm) and a small portion is the 90 degree edge of randomly oriented lipids. The spectrum were acquired with 256 transients at 298 K, employing a Bruker Avance II 14.1 T spectrometer. This sample is the same as **Figure 4.4-B**.

sample) that was reasonably well oriented as judged by ^{2}H **(Figure 4.4-B)** and ^{31}P **(Figure 4.5)** spectra. Unfortunately, I could not see any ^{15}N signal. Potential explanations for this include 1) Estimates of the protein content of the sample were higher than the actual amount of protein in the sample; 2) SP-B conformational heterogeneity/dynamics; 3) Wrong NMR acquisition parameters.

To help diagnose the problem, I ran ^{15}N solid state NMR in different stages of protein purification and preparation, as well as in my mechanically oriented SP-B+POPC/POPG+POPC$_{d31}$ sample, to track the signal intensity and predict the amount of required protein in my sample **(Figure 4.6** and **Table 4.1)**. Here, I

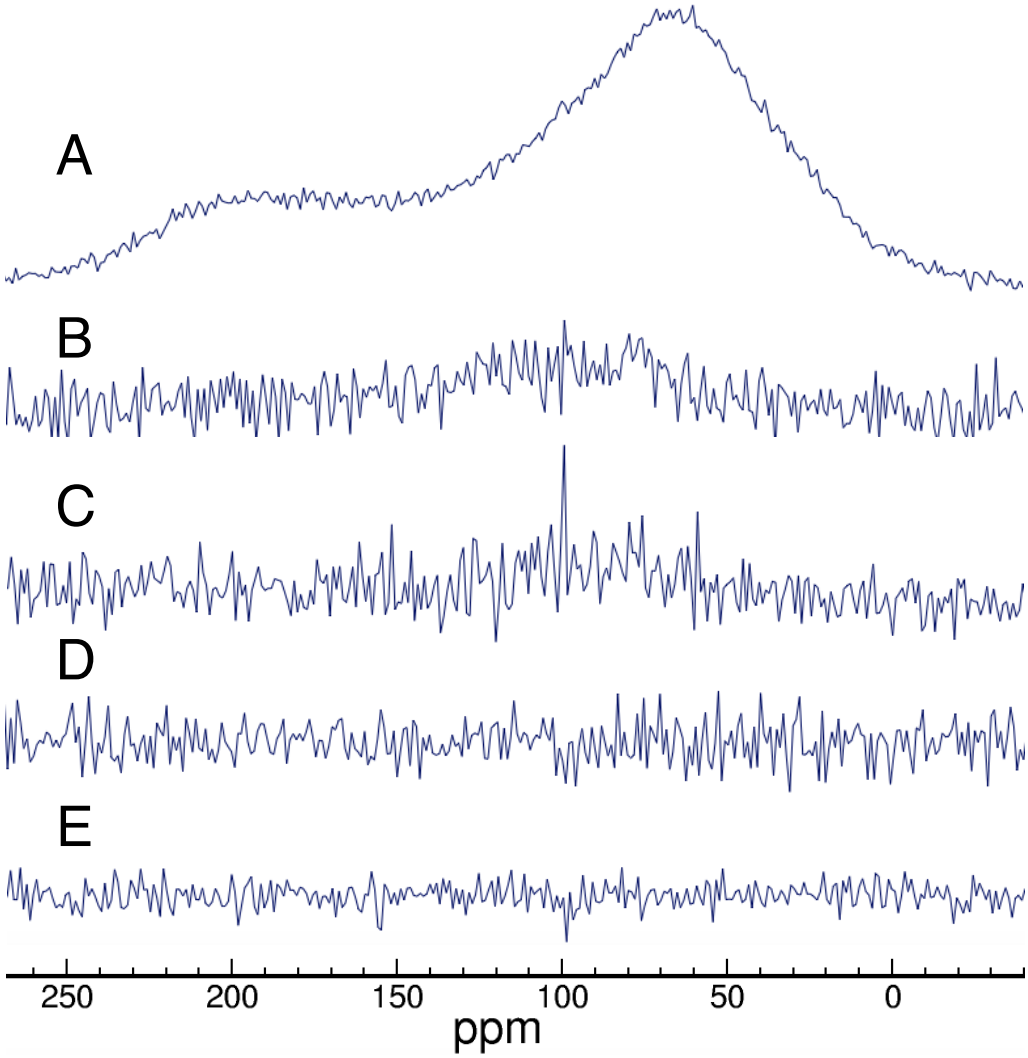

Figure 4.6: The ^{15}N spectra of different steps of SP-B purification and the SP-B+POPC/POPG+POPC$_{d31}$ sample. A) the ^{15}N spectrum of the bacterial lysate, containing the expressed SP-B, as well as bacterial proteins and cellular debris, B) after the first step of purification, C) after the second step of purification, D) the last step of purification, i.e. SP-B+POPC/POPG, E) mechanically oriented SP-B+POPC/POPG+POPC$_{d31}$ sample. We can see that there are no ^{15}N signals at last two steps (D and E). The spectra were acquired with 2900 transients for A, 46300 transients for B, 18400 transients for C, 15700 transients for D, and 110000 transients for E at 298 K, employing a Bruker Avance II 14.1 T spectrometer. The E spectrum is for the same sample as **Figure 4.4-B** and **4.5**. The y-scale is not the same for each spectra.

take a sample at each four steps of protein purification, as well as the final sample in mechanically oriented bilayers, (totally five samples) called: A, B, C, D and E. Sample-A is the lyophilized powder of the bacterial lysate, containing the expressed SP-B, as well as bacterial proteins and cellular debris. Sample-B is the sample we have after the first step of purification. Sample-C is similar to Sample-B, with cleaved proteins and no further purification. Sample-D is the SP-B sample in the last step of purification after desalting the protein, i.e. SP-B+POPC/POPG. Sample-E is the SP-B+POPC/POPG+$POPC_{d31}$ mechanically oriented sample ready to run NMR. All of the samples are in powder form except Sample-E, which contains the SP-B+POPC/POPG+$POPC_{d31}$ bilayers.

We did not measure how much ^{15}N-labelled SP-B was present in samples A to C, which are for SP-B preparation and purification steps. For sample-D and sample-E (which is sample-D+$POPC_{d31}$), based on our previously prepared samples of SP-B in detergents, we expect that the SP-B+POPC+POPG sample (sample-D) contains only 10% of ^{15}N-labelled protein, by weight. In the discussion section we discuss the validity of our estimation.

In **Figure 4.6-A**, we can see the ^{15}N spectrum of 52 mg (visually about 1% of total amount) of sample-A. It takes less than 4 hours of scanning, ∼2900 scans, to get the spectrum, which confirms the high amount of ^{15}N in sample-A **(Figure 4.6-A** and **Table 4.1)**. In my next experiment, I run NMR on 13.3 mg (visually less than 5% of total amount) of sample-B for 46300 scans (about 2 days and 17 hours) **(Figure 4.6-B** and **Table 4.1)**. Here, it appears there might be a small, broad signal above baseline over the 50-150 ppm. The low intensity compared to that of **Figure 4.6-A** is due to the lower amount of ^{15}N nucleus in this sample compared to Sample-A. Next, I run

System	NMR sample mass (mg)	visual % ratio*	number of scans	^{15}N signal
Sample-A	52	1	2900	✓
Sample-B	13.3	5	46300	~✓
Sample-C	84.1	10	18400	~✓
Sample-D	15.4	75	15700	✗
Sample-E	4.8	33	110000	✗

Table 4.1: Measurement summary for each sample, A to E. *The "visual % ratio" indicates the visual % ratio of each sample compared to the the total amount of product, produced in each step.

^{15}N NMR experiment on 84.1 mg (visually about 10% of the total amount) of the sample-C for ~18400 scans, which takes about 1 day **(Figure 4.6-C** and **Table 4.1)**. Here, compared to the sample-B, we run the experiment for a shorter time (1/3 of sample-B), but we have more than 6 times the material by weight. In **Figure 4.6-C** similar to **Figure 4.6-B**, we can see there might be a small, broad signal above baseline over 50-150 ppm. Similar to sample B, in sample C, the lower intensity of signal compared to sample-A is due to the lower amount of ^{15}N nuclei in these samples compared to the sample-A. In **Figure 4.6-D**, we can see the ^{15}N solid state NMR spectrum of the 15.4 mg of the sample-D after 15700 scans **Table 4.1**. We can see that there is no signal in the spectrum, which is due to the very low amount of protein in my sample. In **Figure 4.6-E**, we can see the 1D ^{15}N spectrum of sample-E, i.e. SP-B in the mechanically oriented lipid bilayer. The spectrum has no signal after about 110000 scans (6 days), which could be due to an inadequate amount of the labeled protein in my sample.

4.4 Discussion

The ^{2}H and ^{31}P solid state NMR experiments prove my ability to prepare long lasting mechanically oriented POPC lipid bilayers, using mica plates **(Figure 4.4** and **Figure 4.5)**. The step-by-step ^{15}N solid state NMR experiment indicates that the signal intensity decreases after each step of purification **(Figure 4.6-A** and **Table 4.1)**. In sample-B and sample-C, because of high amounts of lipid added to the sample in the process of purification, the concentration of the labeled protein is low. Thus, we expect a low NMR spectra intensity for these two samples. However, for the final step of SP-B purification, sample-D, we need to modify our dialyzing procedure to remove more lipids from our sample and gain more SP-B as the final product. We may be able to decrease the amount of extra lipids when preparing our stock solution to spread on the plates, however, a larger SP-B to lipid ratio will make it more difficult to produce well oriented bilayers.

In **Figure 4.7**, similar to **Figure 4.6-D**, we can see the ^{15}N NMR signal of 6 mg of dried powder of GAD-1 anti-microbial peptide with 2 of 21 residues selectively ^{15}N-labeled, collected by Gagandeep Sandhu in our group. Here, the signal is obtained at 298 K with 12000 transients, using the same acquisition parameters as in my experiments. This sample is similar to my sample-D, only with different protein contents. A simple calculation indicates that the GAD-1 powder sample (**Figure 4.7**), contains almost 0.6 mg of ^{15}N-labeled residues in the total of 6 mg of GAD-1 peptides:

$$\frac{2\ labeled\ residues}{21\ unlabeled\ residues} \times 6\ mg = 0.57 \simeq 0.6\ mg\ labeled\ residues.$$

Assuming 10% of my sample contains pure ^{15}N-labeled SP-B protein, I should have 1.54 mg of ^{15}N-labeled SP-B (**Table 4.1**), which is a higher amount of labeled protein

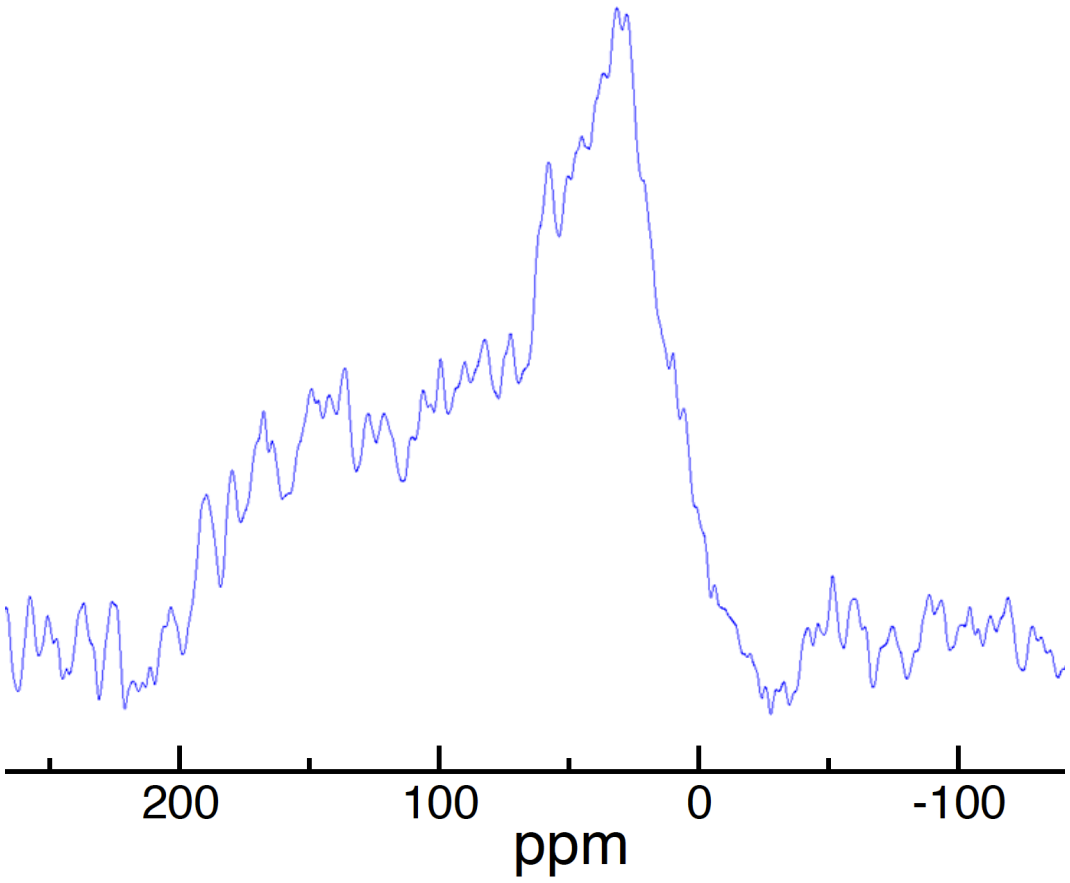

Figure 4.7: The ^{15}N spectra of GAD-1 AMP powder. The spectrum was acquired with 12000 transients at 298 K, employing a Bruker Avance II 14.1 T spectrometer. Used with permission from Gagandeep Sandhu.

compared to Sandhu's sample (**Figure 4.7**).

One possibility for my difficulty in observing an NMR signal might have been with the NMR acquisition parameters used. However, my NMR acquisition parameters, which I used for sample-D in **Figure 4.6-D**, were identical to those used for the spectrum in **Figure 4.7**, so this is not likely to be the case. Another possibility is that my estimate of the weight of the protein in my sample was faulty. This estimate was based on assuming the same protein:lipid ratio in my sample as has been measured for protein:detergent samples. If SP-B binds to much more lipid than it does to detergent, this estimate might be off. However, a sample of SP-B in lipid was sent out for amino acid sequencing by bradford and fluorescamine assays (results not shown here) and the results confirmed our calculated protein concentration. However, this may not exclude the possibility of SP-B binding to the glass vials or pipet tips during sample preparation. One should consider this possibility in the future experiments of SP-B. A third possibility is that SP-B has "conformational flexibility and dynamics" that reduces the signal intensity. Protein conformational flexibility is often a reason for observing less signal than expected in an NMR experiment. However, even in the powder samples, less signal was observed than expected (in sample-D case, no signal). A fourth possibility could be the binding of some ferromagnetic atoms to SP-B during the protein purification. This is an unlikely rare possibility, since the desalting steps in the purification should remove all unwanted salts from the protein.

The all-atom MD simulations in Chapter 3 predict different structures for SP-B interacting with POPC lipid bilayers. However, the predicted ^{15}N spectra of these simulations represent similar plots for these different structures **(Figure 4.3)**. Thus, the uniform ^{15}N labeling of SP-B can not provide us many details of the protein's

orientation in lipid bilayers.

To acquire more details on the orientation and the structure of the protein in lipid bilayers, we could use a residue specific labeling method, e.g. selectively ^{15}N-labelled residues, as well as deuterium-labeled $^{2}H_3$-Ala in the sequence of the protein. In the selectively ^{15}N-labeled method, we grow SP-B on an unlabeled minimal medium with the particular ^{15}N-labeled amino acid(s) of our choice. To employ this method, we can label one or more of Leu, Val or Ala amino acids, which are more frequent residues in the protein's sequence. SP-B contains 14 Leu, 9 Val and 8 Ala residues. Besides the high number of these amino acids in SP-B's sequence, these residues are predicted to be present in the helical portions of the protein, based on the last 200 ns of my MD simulations **(Figure 3.24)**.

In addition to ^{15}N-labelled residues, we can also get structural information from ^{2}H-labelled amino acids. Here, we can add the pitch and tilt angle of the [$^{2}H_3$-Ala] to SP-B, to add more restraints to the orientation of the protein and help us find the final structure. The deuterated alanine(s) should be in the same helical region(s) as the selectively ^{15}N-labelled residues, to perform the information on the orientation of the region with respect to lipid bilayers. The CH_3 group of alanine rotates fast around the $C_\alpha - C_\beta$ bond, which results into a single quadrupole splitting peak, for which the value of the splitting frequency $\Delta\nu_Q$ can be measured and related to peptide orientation via,

$$\Delta\nu_Q = \frac{3}{2}\frac{e^2qQ}{h}\frac{(3\cos^2\Theta - 1)}{2}, \tag{4.5}$$

where $\frac{e^2qQ}{h}$ is the static quadrupole coupling constant and Θ is the angle between the $C_\alpha - C_\beta$ bond and the B_0 [39]. Similar to the ^{15}N-labelled residues, one must use the

PDB structure of the protein to find the orientation of the [$^2H_3$-Ala] in the helical region of the protein. To find the orientation of each helical region of the protein in lipid bilayers, we use the predicted 180×180 counterplot of pitch and tilt angles of ^{15}N-labeled, as well as the ^{2}H-labelled Ala residues in the protein and compare them with the experimental values [30].

In general, similar to the uniformly-labeled SP-B, by using selectively ^{15}N-labelled residues in the sequence of SP-B, if we have a single peak at less than 100 ppm, this would indicate that the whole protein is parallel to the lipid bilayers. Otherwise, if we have a single peak at more than 200 ppm, this would indicate that the whole protein is perpendicular to the lipid bilayers, which is highly unlikely based on my MD results in Chapter 3. However, if we have two or more peaks in the spectra, this could indicate the protein is not entirely parallel to the lipid bilayers.

In addition to 1D ^{15}N experiments, we can use 2D PISEMA experiments, to find the orientation of helical regions of SP-B. In this method, we take the N-H dipolar splitting into account and predict changes in PISEMA plots, while changing the tilt angle of the helical portion of the protein (the PDB file) with respect to the bilayer normal [29, 30]. Then, we compare the predicted results to the experimental results and find the matching tilt angles. In this method, we need clear ^{15}N NMR signals. However, since we have difficulties in acquiring 1D ^{15}N NMR spectrum of SP-B, we can not consider running 2D ^{15}N NMR until we have clear, high-resolution ^{15}N NMR signals of SP-B.

Bibliography

[1] Clark, J. C., Wert, S. E., Bachurski, C. J., Stahlman, M. T., Stripp, B. R., Weaver, T. E., Whitsett, J. A. "Targeted disruption of the surfactant protein B gene disrupts surfactant homeostasis, causing respiratory failure in newborn mice" (1995) Proc Natl Acad Sci U S A 92, 7794-7798.

[2] Hawgood, S., Derrick, M., Poulain, F. "Structure and properties of surfactant protein B" (1998) Biochim Biophys Acta 1408, 150-160.

[3] Munford, R. S., Sheppard, P. O., O'Hara, P. J. "Saposin-like proteins (SAPLIP) carry out diverse functions on a common backbone structure" (1995) J Lipid Res 36, 1653-1663.

[4] Ryan, M. A., Akinbi, H. T., Serrano, A. G., Perez-Gil, J., Wu, H., McCormack, F. X., Weaver, T. E. "Antimicrobial Activity of Native and Synthetic Surfactant Protein B Peptides" (2006) J Immunol. 176(1), 416-425.

[5] Ahn, V. E., Leyko, P., Alattia, J. R., Chen, L. and Prive, G. G. "Crystal structures of saposins A and C" (2006) Protein Sci 15, 1849-1857.

[6] Ahn, V. E., Faull, K. F., Whitelegge, J. P., Fluharty, A. L. and Privé, G. G. "Crystal structure of saposin B reveals a dimeric shell for lipid binding" (2003) Proc Natl Acad Sci USA 100, 38-43.

[7] Anderson, D. H., Sawaya, M. R., Cascio, D., Ernst, W., Modlin, R., Krensky, A. and Eisenberg, D. "Granulysin Crystal Structure and a Structure-derived Lytic Mechanism" (2003) J Mol Biol 325, 355-365.

[8] Hecht, O., Van Nuland, N. A., Schleinkofer, K., Dingley, A. J., Bruhn, H., Leippe, M. and Grötzinger, J. "Solution Structure of the Pore-forming Protein of Entamoeba histolytica" (2004) J Biol Chem 279, 17834-17841.

[9] Liepinsh, E., Andersson, M., Ruysschaert, J. M. and Otting, G. "Saposin fold revealed by the NMR structure of NK-lysin" (1997) Nat Struct Biol 4, 793-795.

[10] de Alba, E., Weiler, S., Tjandra, N. "Solution structure of human saposin C: pH-dependent interaction with phospholipid vesicles" (2003) Biochemistry 42(50), 14729-40.

[11] Sanders, C. R. and Sönnichsen, F. "Solution NMR of membrane proteins: practice and challenges" (2006) Magnetic resonance in chemistry : MRC 44 Spec No, S24-40.

[12] Qureshi, T. and Goto, N. K. "Contemporary Methods in Structure Determination of Membrane Proteins by Solution NMR" (2011) Top Curr Chem 326, 123-185.

[13] Renault, M., Cukkemane, A. and Baldus, M. "Solid-State NMR Spectroscopy on Complex Biomolecules" (2010) Angew Chem Int Ed Engl 49, 8346-8357.

[14] Opella, S. J., Nevzorov, A., Mesleb, M. F. and Marassi, F. M. "Structure determination of membrane proteins by NMR spectroscopy" (2002) Biochem Cell Biol 80, 597-604.

[15] Hong, M. "Structure, Topology, and Dynamics of Membrane Peptides and Proteins from Solid-State NMR Spectroscopy" (2007) The journal of physical chemistry B 111, 10340-10351.

[16] McDermott, A. "Structure and dynamics of membrane proteins by magic angle spinning solid-state NMR" (2009) Annu Rev Biophys 38, 385-403.

[17] Sarker, M., Waring, A. J., Walther, F. J., Keough, K. M. W. and Booth, V. "Structure of Mini-B, a Functional Fragment of Surfactant Protein B, in Detergent Micelles" (2007) Biochemistry 46, 11047-11056.

[18] Sharifahmadian, M., Sarker, M., Palleboina, D., Waring, A. J., Walther, F. J., Morrow, M. R., Booth, V. "Role of the N-Terminal Seven Residues of Surfactant Protein B (SP-B)" (2013) PLoS ONE 8(9), e72821.

[19] Booth, V., Waring, A. J., Walther, F. J., Keough, K. M. "NMR structures of the C-terminal segment of surfactant protein B in detergent micelles and hexafluoro-2-propanol" (2004) Biochemistry. 43(48),15187-94.

[20] Perez-Gil. J., Cruz. A., Casals. C. "Solubility of hydrophobic surfactant proteins in organic solvent/water mixtures. Structural studies on SP-B and SP-C in

aqueous organic solvents and lipids" (1993) Biochim Biophys Acta 1168, 261-270.

[21] Vandenbussche, G., Clercx, A., Clercx, M., Curstedt, T., Johansson, J., Jornvall, H., Ruysschaert, J. M. "Secondary structure and orientation of the surfactant protein SP-B in a lipid environment. A Fourier transform infrared spectroscopy study" (1992) Biochemistry 31, 9169-9176.

[22] Andersson, M., Curstedt, T., Jornvall, H., Johansson, J. "An amphipathic helical motif common to tumourolytic polypeptide NK-lysin and pulmonary surfactant polypeptide SP-B" (1995) FEBS Lett. 362, 328-332.

[23] Bertani, P., Vidovic, V., Yang, T., Rendell, J., Gordon, L. M., Waring, A. J., Bechinger, B., Booth, V. "Orientation and depth of surfactant protein B C-terminal helix in lung surfactant bilayers" (2012) Biochimica et Biophysica Acta 1818, 1165-1172.

[24] Yang, T., McDonald, M., Morrow,M. R., Booth, V. "The Effect of a C-Terminal Peptide of Surfactant Protein B (SP-B) on Oriented Lipid Bilayers, Characterized by Solid-State ^{2}H- and ^{31}P-NMR" (2009) Biophys J. 96(9), 3762-3771.

[25] Bruhn, H. "A short guided tour through functional and structural features of saposin-like proteins" (2005) Biochem J 389, 249-257.

[26] Olmeda, B., García-Aĺvarez, B., Peŕez-Gil, J. "Structure-function correlations of pulmonary surfactant protein SP-B and the saposin-like family of proteins" (2013) Eur Biophys J 42, 209-222.

[27] Olmeda, B., Garcia-Alvarez, B., Gomez, M. J., Martinez-Calle, M., Cruz, A. and Perez-Gil, J. "A model for the structure and mechanism of action of pulmonary surfactant protein B" (2015) The FASEB Journal 29(10) 4236-4247.

[28] Kandasamy, S. K., Lee, D. K., Nanga, R. P., Xu, J., Santos, J. S., Larson, R. G., Ramamoorthy, A. "Solid-state NMR and molecular dynamics simulations reveal the oligomeric ion-channels of TM2-GABA(A) stabilized by intermolecular hydrogen bonding" (2009) Biochim Biophys Acta. 1788(3), 686-95.

[29] Kawaguchi, K., Suita, K., Suzuki, Y., Umemoto, K. Nakazawa, Y., Asakura T. "Orientation of the Antimicrobial Peptide, Cecropin AÐMagainin 2 Hybrid, in a Lipid Bilayer Studied by ^{15}N Solid-State NMR" (2005) Polymer Journal 37(3), 229-233.

[30] Aisenbrey, C., Sudheendra, U. S., Ridley, H., Bertani, P., Marquette, A., Nedelkina, S., Lakey, J. H. and Bechinger, B. "Helix orientations in membrane-associated Bcl-X(L) determined by ^{15}N-solid-state NMR spectroscopy" (2007) Eur Biophys J 37, 71-80.

[31] Vosegaard, T., Kamihira-Ishijima, M., Watts, A. and Nielsen, N. C. "Helix Conformations in 7TM Membrane Proteins Determined Using Oriented-Sample Solid-State NMR with Multiple Residue-Specific ^{15}N Labeling" (2008) Biophysical Journal 94, 241-250.

[32] Rainey, J., K., and Sykes, B., D. "Optimizing oriented planar-supported lipid samples for solid-state protein NMR" (2005) Biophysical Journal 89, 2792-2805.

[33] Davis, J. H., Jeffrey, K. R., Bloom, M., Valic, M. I., Higgs, T. P. "Quadrupolar echo deuteron magnetic resonance spectroscopy in ordered hydrocarbon chains" (1976) Chem Phy Lett 42 (2), 390-394.

[34] Hartzell, C. J., Whitfield, M., Oas, T. G., Drobny, G. P. "Determination of the ^{15}N and ^{13}C chemical shift tensors of L-[^{13}C]alanyl-L-[^{15}N]alanine from the dipole-coupled powder patterns" (1987) J Am Chem Soc 109, 5966-5969.

[35] Lee, D. K., Wittebort, R. J., Ramamoorthy, A. "Characterization of ^{15}N Chemical Shift and ^{1}H?^{15}N Dipolar Coupling Interactions in a Peptide Bond of Uniaxially Oriented and Polycrystalline Samples by One-Dimensional Dipolar Chemical Shift Solid-State NMR Spectroscopy" (1998) J Am Chem Soc 120, 8868-8874.

[36] Lee, D.K., Wei, Y., Ramamoorthy, A. "A Two-Dimensional Magic-Angle Decoupling and Magic-Angle Turning Solid-State NMR Method:? An Application to Study Chemical Shift Tensors from Peptides That Are Nonselectively Labeled with ^{15}N Isotope" (2001) J Phys Chem B 105, 4752-4762.

[37] Salnikov, E., Aisenbrey, C., Vidovic, V., Bechinger, B. "Solid-state NMR approaches to measure topological equilibria and dynamics of membrane polypeptides" (2010) Biochimica et Biophysica Acta 1798, 258-265.

[38] Bechinger, B., Resende, J. M., Aisenbrey, C. "The structural and topological analysis of membrane-associated polypeptides by oriented solid-state NMR spectroscopy: Established concepts and novel developments" (2011) Biophysical Chemistry 153, 115-125.

[39] Bechinger, B., Aisenbrey, C., Bertani, P. "The alignment, structure and dynamics of membrane-associated polypeptides by solid-state NMR spectroscopy" (2004) Biochimica et Biophysica Acta 1666, 190-204.

[40] Bechinger, B. and Sizun, C. "Alignment and structural analysis of membrane polypeptides by ^{15}N and ^{31}P solid-state NMR spectroscopy" (2003) Concepts Magn. Reson. 18A, 130-145.

[41] Esteban-Martin, S., Strandberg, E., Fuertes, G., Ulrich, A. S., Salgado, J. "Influence of Whole-Body Dynamics on ^{15}N PISEMA NMR Spectra of Membrane Proteins: A Theoretical Analysis" (2009) Biophysical Journal 96, 3233-3241.

[42] Frishman, D., Argos, P. "Knowledge-based protein secondary structure assignment " (1995) Proteins: Structure, Function, and Genetics 23, 566-579.

[43] Saito, H., Ando, I., Naito, A. " Solid State NMR Spectroscopy for Biopolymers: Principles and applications" (2006) Springer, chapter 14, page 376.

Chapter 5

Conclusions and future work

During my PhD studies, I employed computational simulations and solid state NMR experiments to study two types of membrane-active proteins, antimicrobial peptides (AMPs) and lung surfactant (LS) proteins. Chapter 1 provides a short, brief introduction to definitions and methods I use in my thesis. In the first part of my studies (discussed in Chapter 2), I employed molecular dynamics simulations to study two histidine-containing antimicrobial peptide paralogs (GAD-1 and GAD-2) derived from cod, interacting with self-assembled bilayers. This work is published in BBA-Biomembranes journal [1]. In the second part of my studies, I worked on computational (discussed in Chapter 3) and experimental (discussed in Chapter 4) studies of pulmonary surfactant protein B interacting with lipid bilayers. In this chapter, I provide a short summary of my studies in each chapter of my thesis along with ideas for future work. A shorter version of chapter 3 (mainly the all-atom force field simulations) is accepted to be published as an article in BBA-Biomembranes journal [2].

5.1 Computational Simulations of AMPs

In my first project, I worked on GAD-1 and GAD-2 antimicrobial peptide paralogs derived from codfish [3–6]. Since our peptides were rich in histidine (which is pH sensitive), we used our peptides with both neutral and positively charged histidines. In our MD simulations, following the method used by Salgado et al. [7] we used a self-assembly method to set up our systems. Here, most systems formed a bilayer with a stable pore in it, which lasted during the simulation time. The bilayer self-assembled in the presence of a peptide, and thus the peptide could find low energy conformations within the membrane without being biased by the initial system configuration. Thus, the study did not address the steps involved in peptide-induced pore formation in a well-formed membrane. The most interesting result of our simulations stemmed from the observation that the N-terminus half of GAD peptides tended to interact with the pore region of the lipid bilayer. We indicated that this preference was related to the position of histidine pairs in the sequence of peptides, not the other types of charged pairs, e.g. histidine-arginine. In this work, we discussed that due to the constrained structure of histidines (regardless of whether the histidine is neutral or positively charged), the more aqueous pore regions provided an energetically more favourable site for the histidine pairs.

This work could be extended by running the simulations much longer for the protein structures to converge and reach an equilibrium [8]. Future studies could include simulating the GAD peptides in an aqueous environment, i.e., without lipids. This will provide another point of comparison of structure with experiment, and may be important as AMP's are transported throughout the body via the blood.

The results could be compared to the ones experimentaly studied by McDonald, et al. [9]. To explore the peptide-induced pore formation steps, rather than using self-assembly method employed in chapter 2, we could run new sets of simulations, and place peptides inside, half inside and outside a pre-formed lipid bilayer, similar to my simulations in Chapter 3. In order to study GADs in multimeric structures, we can run our systems with more than a single peptide in each simulation box, similar to Mihajlovic et al. [10], in which they used two pre-assembled toroidal and cylindrical pores and place their proteins inside the pore. They indicated that the type of the pore strongly depends on the charge of the protein. In this way, we could study the ability of our antimicrobial peptides to form multimers, as well as importance of pH change in the change of pore preference. Simulating both GAD-1 and GAD-2 peptides together could provide insight on whether and how these two peptides could function synergistically.

In another study, Kim et al. [11], used gramicidin A antimicrobial peptide in monomeric and dimeric structures to study the pore formation in different lipid types. They studied the influence of hydrophobic mismatch on the pore type. Similarly, we could use more complex lipid mixtures, e.g. by adding negatively-charged lipid molecules like PG (which is more frequent in bacterial membranes) and other lipid types to study the importance of hydrophobic mismatch and charge preference of our AMPs. In addition to placing our peptides inside the lipid bilayer, we can put them randomly in the simulation box, similarly to Leontiadou et al. [12], where they studied the effect of magainin MG-H2 peptide aggregation on pore formation. In their studies they placed their peptide randomly near the bilayer surface. After running their simulations, peptides bound to the bilayer and aggregated to form a

pore, while the non-aggregated peptides remained parallel to the membrane. As another suggestion, we can compare our computational results with the solid state NMR results currently being performed by Gagandeep Sandhu in our group, similar to the work of Pino-Angeles et al. [8] and Thogersen et al. [13]. In these studies, they run long time simulations of multimers of AMPs in pre-assembled bilayers and compare their structure with the experimental ones.

To use other methods of simulation, in addition to replica exchange MD simulations (which is explained in Chapter 3) we can use Coarse-Grained (CG) force fields to run our simulations for a long time and convert the equilibrated system to all-atom simulation, similar to Thogersen et al. [13], in which they run their simulations in CG and convert it to all-atom to run for 50 ns. To characterize the free energy of pore formation, we could use umbrella sampling method similar to the work of Mirjalili et. al. [14].

5.2 Computational Simulations of SP-B

In Chapter 3, I worked on computational studies of pulmonary surfactant protein SP-B interacting with lipid bilayers. Here, I used MD (employing OPLS-AA [15, 16] and PACE [17, 18] force fields) and REMD (employing PACE force field) methods to run my simulations. Since the 3D structure of SP-B is unknown, I constructed my model of SP-B, based on the structure of Mini-B (construct of SP-B with known experimental structure) with the help of homology models of NK-lysin [19] and Saposin-C [20,21] and theoretical predictions on the structure of SP-B. For each simulation, I used a monomer of SP-B in open and V-shaped bent structures, initially placed inside, half

inside and outside the POPC lipid bilayer. Most of the OPLS-AA MD simulations ran for 2–2.5 μs while all MD and REMD simulations using PACE force field ran for 3 μs.

Our all-atom MD simulations provides us with energetically feasible structures of SP-B within or near bilayer. Besides, final structures fine-tune the secondary structure predicted by our model. The interesting result in our simulation was that the central loop and R52 region can help promote non-lamellar lipid structures and structural plasticity and salt bridges play an important role in SP-B/lipid interactions. In our PACE simulations (MD and REMD), proteins are mainly parallel to the lipid bilayer. In these simulations, the proteins lack the plasticity in the structure, which prevents them form producing bent structure. In these PACE simulations, the first 7 residues lack the helicity structure. Despite the similarity in the results of the MD and REMD simulations, REMD simulations provide transmembrane structures, which are not supported by NMR experiments [22].

One thing that the PACE simulations showed was the unwinding of the helicity results indicated of first 7 residues in the protein. We can improve our initial model of protein, using results obtained in my simulations, e.g. remove helicity from the ERYS sequence in helix-IV, or even use other hypotheses in the literature, e.g. construct the first seven residues of the protein N-terminal in a random coil structure. Moreover, to enhance the setup procedure, in addition to the open and bent structures, we can use a closed structure as an initial structure for SP-B. For these simulations, we can obtain our initial structure from a closed structure, formed at the end of the open-out and/or bent-out simulations and place it inside, half inside and outside the lipid bilayer. Besides, considering that SP-B may be functional in dimeric and/or

multimeric structures, our new simulations could contain dimer(s) of SP-B similar to Baoukina et al.'s coarse-grained simulations [23], where they studied the role of SP-B monomers in the fusion of two lipid vesicles, in the presence of SP-C proteins. In their simulations they randomly placed the proteins in their system and ran it for > 5 μs. Although their simulations lacked the details of SP-B's role owing to their use of the MARTINI coarse-grained force field [24–26], their results indicated that the presence of SP-B facilitates the the fusion of vesicles with each other. Moreover, to mimic the lung surfactant lipids in our simulations better, we can use more complex lipid mixtures by adding DPPC and POPG lipid molecules, which are present in the lung lipid contents. We can also produce more complex lipid structures by adding monolayers of lipid into our simulation system, similar to Baoukina et al. [27], where they studied diffusion of two lipid vesicles and a lipid bicelle into lipid monomers in the presence of SP-B monomers and dimers. After being familiar with the system, we can run our simulations using bilayer self-assembly methods, similar to the works in Chapter 2. Depending on the system size, self-assembly method could consume a high amount of simulation time. However, this method could provide information on more favoured environments for SP-B within or near a lipid bilayer.

5.3 Solid State NMR Experiments on SP-B

In Chapter 4, I followed up on my studies of SP-B interacting with lipid bilayers. Here, instead of using computational simulations, I used ^{15}N, ^{2}H and ^{31}P solid-state NMR experiments to study my system. In Chapter 4, the goal was to compare the experimental NMR spectrum with the computational results from Chapter 3. I

used an ^{15}N-labelled SP-B sample, which was recombinantly expressed in our group. Unfortunately, I was unable to obtain any ^{15}N NMR signal from my oriented protein+lipid sample, most probably owing to the insufficient amount of protein in my NMR samples. Although I could not reach the final goal of this chapter, my work presented information that could be helpful for similar systems. In order to compare with experiments, I calculated ^{15}N NMR spectra of SP-B, based on the computationally predicted SP-B structures of chapter 3 using the method introduced in ref. [28] (the procedure is described in Chapter 4 and Appendix B).

In addition to improving the efficiency of SP-B purification, to get more reliable results in future experiments, we should use more complex lipid mixtures in our samples, e.g. by adding DPPC and POPG lipid molecules, which mimic the LS lipid contents better. The computational structures could provide information on how to specifically label SP-B in order to gain detailed structure of my system. Here, we can label residues, which are predicted to be present in a specific helical region and see if the helix exists. The code I employed to predict NMR spectra from my simulations could be used for other proteins, e.g. GAD-1 and GAD-2. This could help reveal information on the orientation of GAD peptides, currently being studied by Gagandeep Sandhu employing solid-state NMR. Here, one could construct a predicted structure of GADs and rotate it in different orientations with respect to lipid bilayer to calculate the NMR spectrum.

Bibliography

[1] Khatami, M. H., Bromberek, M., Saika-Voivod, I., Booth, V. "Molecular dynamics simulations of histidine-containing cod antimicrobial peptide paralogs in self-assembled bilayers" (2014) BBA - Biomembranes 1838, 2778-2787.

[2] Khatami, M. H., Saika-Voivod, I., Booth, V. "All-atom Molecular Dynamics Simulations of Lung Surfactant Protein B: Structural Features of SP-B Promote Lipid Reorganization" (in press 2016) BBA - Biomembranes.

[3] Browne, M. J., Feng, C. Y., Booth, V. and Rise, M. L. "Characterization and expression studies of Gaduscidin-1 and Gaduscidin-2; paralogous antimicrobial peptide-like transcripts from Atlantic cod (Gadus morhua)" (2011) Dev Comp Immunol 35, 399-408.

[4] Ruangsri, J., Fernandes, J. M. O., Rombout, J. H. W. M., Brinchmann, M. F. and Kiron, V. "Ubiquitous presence of piscidin-1 in Atlantic cod as evidenced by immunolocalisation" (2012) BMC veterinary research 8, 46.

[5] Ruangsri, J., Salger, S. A., Caipang, C., Kiron, V. and Fernandes, J. M. O. "Differential expression and biological activity of two piscidin paralogues and a

novel splice variant in Atlantic cod (Gadus morhua L.)" (2012) Fish & Shellfish Immunology 32, 396-406.

[6] Fernandes, J. M. O., Ruangsri, J. and Kiron, V. "Atlantic Cod Piscidin and Its Diversification through Positive Selection" (2010) Plos one 5, e9501.

[7] Esteban-Martín, S. and Salgado, J. "Self-Assembling of Peptide/Membrane Complexes by Atomistic Molecular Dynamics Simulations" (2007) Biophysical journal 92, 903-912.

[8] Pino-Angeles, A., Leveritt III, J. M., Lazaridis, T. "Pore Structure and Synergy in Antimicrobial Peptides of the Magainin Family" (2016) PLoS Comput Biol 12(1): e1004570.

[9] McDonald, M., Mannion, M., Pike, D., Lewis, K., Flynn, A., Brannan, A. M., Browne, M. J., Jackman, D., Madera, L., Power Coombs, M. R., Hoskin, D. W., Rise, M. L., Booth, V. "Structure-function relationships in histidine-rich antimicrobial peptides from Atlantic cod" (2015) Biochimica et Biophysica Acta (BBA) - Biomembranes 1848(7), 1451-1461.

[10] Mihajlovic, M., Lazaridis, T. "Antimicrobial peptides in toroidal and cylindrical pores" (2010) Biochimica et Biophysica Acta 1798,1485-1493.

[11] Kim, T., Lee, K. I., Morris, P., Pastor, R. W., Andersen, O. S., Im, W. "Influence of Hydrophobic Mismatch on Structures and Dynamics of Gramicidin A and Lipid Bilayers" (2012) Biophysical Journal 102(7), 1551-1560.

[12] Leontiadou, H., Mark, A. E., Marrink, S. J. "Antimicrobial Peptides in Action" (2006) J. AM. CHEM. SOC. 128(37), 12156-12161.

[13] Thogersen, L., Schiott, B., Vosegaard, T., Nielsen, N., C., Tajkhorshid, E. "Peptide Aggregation and Pore Formation in a Lipid Bilayer: A Combined Coarse-Grained and All Atom Molecular Dynamics Study" (2008) Biophysical Journal 95(5) 4337-4347.

[14] Mirjalili, V., Feig, M. "Density-Biased Sampling: A Robust Computational Method for Studying Pore Formation in Membranes" (2015) J. Chem. Theory Comput. 11, 343-350.

[15] Jorgensen, W. L., Maxwell, D. S., Tirado-Rives, J. "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids" (1996) Journal of the American Chemical Society 118, 11225-11236.

[16] Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., Jorgensen, W. L. "Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides" (2001) J. Phys. Chem. B 105, 6474-6487.

[17] Han, W., Schulten, K. "Further Optimization of a Hybrid United-Atom and Coarse-Grained Force Field for Folding Simulations: Improved Backbone Hydration and Interactions between Charged Side Chains" (2012) J. Chem. Theory Comput. 8, 4413-4424.

[18] Wan, C. K., Han, W., Wu, Y. D. "Parameterization of PACE Force Field for Membrane Environment and Simulation of Helical Peptides and Helix-Helix Association" (2012) J. Chem. Theory Comput. 8, 300-313.

[19] Liepinsh, E., Andersson, M., Ruysschaert, J. M. and Otting, G. "Saposin fold revealed by the NMR structure of NK-lysin" (1997) Nat Struct Biol 4, 793-795.

[20] Hawkins, C. A., de Alba, E. and Tjandra, N. "Solution structure of human saposin C in a detergent environment" (2005) J Mol Biol 346, 1381-1392.

[21] de Alba, E., Weiler, S. and Tjandra, N. "Solution Structure of Human Saposin C: pH-Dependent Interaction with Phospholipid Vesicles" (2003) Biochemistry 42, 14729-14740.

[22] Sharifahmadian, M., Sarker, M., Palleboina, D., Waring, A. J., Walther, F. J., Morrow, M. R., Booth, V. "Role of the N-Terminal Seven Residues of Surfactant Protein B (SP-B)" (2013) PLoS ONE 8(9), e72821.

[23] Baoukina, S. and D. P., Tieleman "Direct Simulation of Protein-Mediated Vesicle Fusion: Lung Surfactant Protein B" (2010) Biophysical Journal 99(7) 2134-2142.

[24] Marrink, S. J., De Vries, A. H., Mark, A. E. J. "Coarse Grained Model for Semiquantitative Lipid Simulations" (2004) Phys. Chem. B 108, 750-760.

[25] Marrink, S. J., Risselada, H. J., Yefimov, S., , D. P., De Vries, A. H. "The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations" (2007) J. Phys. Chem. B 111, 7812-7824.

[26] Monticelli, L., Kandasamy, S. K., Periole, X., Larson, R. G., Tieleman, D. P., Marrink, S. J. "The MARTINI Coarse-Grained Force Field: Extension to Proteins" (2008) J. Chem. Theory Comput. 4, 819-834.

[27] Baoukina, S. and D. P., Tieleman "Lung Surfactant Protein SP-B Promotes Formation of Bilayer Reservoirs from Monolayer and Lipid Transfer between the Interface and Subphase" (2011) Biophysical Journal 100(7) 1678-1687.

[28] Bechinger, B. and Sizun, C. "Alignment and structural analysis of membrane polypeptides by ^{15}N and ^{31}P solid-state NMR spectroscopy" (2003) Concepts Magn. Reson. 18A, 130-145.

[29] Kandasamy, S. K., Lee, D. K., Nanga, R. P., Xu, J., Santos, J. S., Larson, R. G., Ramamoorthy, A. "Solid-state NMR and molecular dynamics simulations reveal the oligomeric ion-channels of TM2-GABA(A) stabilized by intermolecular hydrogen bonding" (2009) Biochim Biophys Acta. 1788(3), 686-95.

[30] Bechinger, B., Resende, J. M., Aisenbrey, C. "The structural and topological analysis of membrane-associated polypeptides by oriented solid-state NMR spectroscopy: Established concepts and novel developments" (2011) Biophysical Chemistry 153, 115-125.

Appendix A

GROMACS input files

To run MD simulations using GROMACS, we should prepare the simulation box by energy minimization, NVT and finally NPT steps. In the first step, we should apply a force field of our choice to the protein, using pdb2gmx command. This force field could be one of the default force fields of GROMACS or a new (or customized) force field, which you have installed in the force fields folder (if you use the default path way in GROMACS installation it should be in: /usr/local/gromacs/share/gromacs/top). After this step, GROMACS makes a ".itp" file, which contains the force field details applied to our system, a ".top" file, which contains the topology of details of our system and a ".gro" file, which contains details of system size and the position of each particle in our system. If our system contains lipid bilayers, we should manually add the corresponding files of the bilayer to the .gro, .top and .itp files. After adding the bilayer, we should add solvent and ions to our system. In this step, the whole content of the simulation box is complete and we are ready to run energy minimize the system. To run MD simulations, GROMACS need a ".mdp" file, which contains details

of simulation like running parameters, output control, bond parameters, neighbour searching methods, electrostatics, temperature and pressure couplings.

An example of a .top file, while the system is ready to run MD simulation is as follow:

; Include forcefield parameters

#include “oplsaa.ff/forcefield.itp”

#include “lipid_se_oplsaa_LM_Nov06.itp”

#include “POPC_LM_Nov2006.itp”

#include “spb.itp”

; Include Position restraint file

#ifdef POSRES

#include “spb_porse.itp”

#endif

; Include water topology

#include “oplsaa.ff/tip4p.itp”

#ifdef POSRES_WATER

; Position restraint for each water oxygen

[position_restraints]

; i funct fcx fcy fcz

1 1 1000 1000 1000

#endif

; Include topology for ions

#include “oplsaa.ff/ions.itp”

[system]

; Name

SP-B and POPC in water

[molecules]

; Compound

Protein 1

POPC 496

SOL 38000

CL 7

An example of a .mdp file, while the system is ready to run NPT simulation is as follow:

define = -DFLEXIBLE ; position restrain the protein

include = -I../include ; include directory with itp files

; Run parameters

integrator = md ; leap-frog integrator

nsteps = 3000000000 ; 0.001 * 500000 = 500 ps

dt = 0.002 ; 2 fs

; Output control

nstxout = 500000 ; save coordinates every 0.5 ps

nstvout = 500000 ; save velocities every 0.5 ps

nstenergy = 500000 ; save energies every 0.5 ps

nstlog = 500000 ; update log file every 0.5 ps

; Bond parameters

continuation = no ; Restarting after NVT

```
constraint_algorithm = lincs              ; holonomic constraints
constraints = all-bonds              ; all bonds (even heavy atom-H bonds) constrained
lincs_iter = 1              ; accuracy of LINCS
lincs_order = 3              ; also related to accuracy
    ; Neighborsearching
ns_type = grid              ; search neighboring grid cels
nstlist = 5              ; 10 fs
rlist = 1.2              ; short-range neighborlist cutoff (in nm)
rcoulomb = 1.2              ; short-range electrostatic cutoff (in nm)
rvdw = 1.2              ; short-range van der Waals cutoff (in nm)
    ; Electrostatics
coulombtype = PME              ; Particle Mesh Ewald for long-range electrostatics
pme_order = 4              ; cubic interpolation
fourierspacing = 0.15              ; grid spacing for FFT
    ; Temperature coupling is on
tcoupl = Nose-Hoover              ; More accurate thermostat
tc-grps = Protein POPC CL SOL              ; three coupling groups - more accurate
tau_t = 0.1 0.1 0.1 0.1              ; time constant, in ps
ref_t = 310 310 310 310              ; reference temperature, one for each group, in K
    ; Pressure coupling is on
pcoupl = Parrinello-Rahman              ; Pressure coupling on in NPT
pcoupltype = semiisotropic              ; uniform scaling of x-y-z box vectors
tau_p = 5.0              ; time constant, in ps
ref_p = 1 1              ; reference pressure, x-y, z (in bar)
```

```
compressibility = 4.5e-5 4.5e-5          ; isothermal compressibility, bar^-1
    ; Periodic boundary conditions
pbc = xyz              ; 3-D PBC
    ; Selection of energy groups =
energygrps = Protein POPC CL SOL


    ; Dispersion correction
DispCorr = EnerPres            ; account for cut-off vdW scheme
    ; Velocity generation
gen_vel = yes            ; Velocity generation is off
gen_temp = 310             ; temperature for Maxwell distribution
gen_seed = -1            ; generate a random seed
; COM motion removal
; These options remove motion of the protein/bilayer relative to the solvent/ions
nstcomm = 1
comm-mode = Linear
comm-grps = Protein POPC CL SOL
```

Appendix B

Mathematica calculations

I used MATHEMATICA v.9 to write my code based on **Equation 4.2** in Chapter 3. The code is as follows (comments are in italics inside parentheses below each line of code):

```
Remove["Global`*"]
```

*(*This line removes all the previously saved symbols in the MATHEMATICA code, which allows us to reevaluate of the entire code, while run it each time. *)*

```
SetDirectory[“Directory name”]
```

*(*This line sets your input files directory to be reached. *)*

```
FN = FileNames[“N–H_*”];
```

*(*This line calls the N–H input files as “FN”. *)*

```
alldata = Map[Import[#, “Table”] &, FN];
```

*(*This line imports “FN” data in “alldata” for each time steps. *)*

Do[Do[$\cos\theta_{i,j}$ =

alldata[[i, j,3]]$\Big/\sqrt{(alldata[[i,j,1]]^2 + alldata[[i,j,2]]^2 + alldata[[i,j,3]]^2)}$,

{j, 1, Length[alldata[[i]]], 1}], {i, 1, Length[FN], 1}];

*(*This line calculates the* $\cos\Theta$ *values of N–H vectors as described in* **Figure 4.1-A**.**)*

value = {65, 85, 230};

Do[Do[$\sigma 11_{i,j}$ = value [[1]], {j, 1, Length[alldata[[i]]], 1}], {i, 1, Length[FN], 1}];

Do[Do[$\sigma 22_{i,j}$ = value [[2]], {j, 1, Length[alldata[[i]]], 1}], {i, 1, Length[FN], 1}];

Do[Do[$\sigma 33_{i,j}$ = value [[3]], {j, 1, Length[alldata[[i]]], 1}], {i, 1, Length[FN], 1}];

*(*These lines inputs the* σ_{11}, σ_{22} *and* σ_{33} *values respectively.*)*

Do[Do[σparallel$_{i,j}$ = $\sigma 33_{i,j}$, {j, 1, Length[alldata[[i]]], 1}], {i, 1, Length[FN], 1}];

*(*This line calculates* $\sigma_{\|}$ *based on* **Equation 4.2**.**)*

Do[Do[σperpendicular$_{i,j}$ = $\frac{1}{2}$*($\sigma 11_{i,j}$+$\sigma 22_{i,j}$)), {j, 1, Length[alldata[[i]]], 1}], {i, 1, Length[FN], 1}];

*(*This line calculates* $\sigma_{\perp}$ *based on* **Equation 4.2**.**)*

Do[Do[σzz$_{i,j}$ = (σparallel$_{i,j}$–σperpendicular$_{i,j}$)*$\cos^2\theta$+ σperpendicular$_{i,j}$, {j, 1, Length[alldata[[i]]], 1}], {i, 1, Length[FN], 1}];

*(*This line calculates* σ_{zz} *based on* **Equation 4.2**.**)*

Do[Do[file$_{i,j}$ = NormalDistribution [σzz$_{i,j}$, 2.5],

{j, 1, Length[alldata[[i]]], 1}], {i, 1, Length[FN], 1}];

*(*This lines applies a normal distribution with standard deviation of 2.5 to the σ_{zz} values.*)*

reverseplot= ListLinePlot[Table[Sum[Sum[PDF[$file_{i,j}$, x], {j, 1, Length[alldata[[i]]], 1}], {i,1, Length[FN], 1}], {x, 0, 250}], ScalingFunctions → {"Reverse", Identity},

TicksStyle → Directive[FontSize → 28], Axes →{True, False},

AxesLabel → {Style["ppm", Black, FontSize→ 32], " "},

PlotRange → {{0, 240}, {0, 30}}, ImageSize →Large]

*(*This line plots the σ_{zz} spetra for all the N–H input values for all time steps.*)*

Export["spectra.pdf", reverseplot]

*(*This line exports the σ_{zz} (^{15}N) spectra plot as a pdf file.*)*